

Design of PAR-Constrained Sequences for MIMO Channel Estimation via Majorization-Minimization

Zhongju Wang, Prabhu Babu, and Daniel P. Palomar, *Fellow, IEEE*

Abstract—PAR-constrained sequences are widely used in communication systems and radars due to various practical needs; specifically, sequences are required to be unimodular or of low peak-to-average power ratio (PAR). For unimodular sequence design, plenty of efforts have been devoted to obtaining good correlation properties. Regarding channel estimation, however, sequences of such properties do not necessarily help produce optimal estimates. Tailored unimodular sequences for the specific criterion concerned are desirable especially when the prior knowledge of the channel is taken into account as well. In this paper, we formulate the problem of optimal unimodular sequence design for minimum mean square error estimation of the channel impulse response and conditional mutual information maximization, respectively. Efficient algorithms based on the majorization-minimization framework are proposed for both problems with guaranteed convergence. As the unimodular constraint is a special case of the low PAR constraint, optimal sequences of low PAR are also considered. Numerical examples are provided to show the performance of the proposed training sequences, with the efficiency of the derived algorithms demonstrated.

Index Terms—Unimodular sequence, peak-to-average power ratio (PAR), channel estimation, majorization-minimization, minimum mean square error, conditional mutual information.

I. INTRODUCTION

PAR-CONSTRAINED sequences, such as unimodular or low peak-to-average power ratio (PAR), have many applications in both single-input single-output (SISO) and multi-input multi-output (MIMO) communication systems. For example, the M -ary phase-shift keying techniques allow only symbols of constant-modulus, i.e., unimodular, to be transmitted [1]. In MIMO radars and code-division multiple-access (CDMA) applications, the practical implementation demands from hardware, such as radio frequency power amplifiers and analog-to-digital converters, require the sequences transmitted to be unimodular or low PAR [2]–[4]. In this paper, we consider the design of optimal unimodular or low PAR sequences for channel estimation.

There is an extensive literature on designing single unimodular sequences with good correlation properties such that the autocorrelation of the sequence is zero at each nonzero lag. As such properties are usually difficult to achieve, metrics of “goodness” have been proposed instead where autocorrelation sidelobes are suppressed rather than literally set to zero, and

optimization problems are thus formulated and solved with numerical algorithms [5], [6]. Specifically, the work [5] provides several cyclic algorithms (CA) for either minimizing integrated sidelobe level (ISL) or maximizing ISL-related merit factor (MF). In [6], a computationally efficient algorithm called MISL for minimizing ISL is proposed, and it is demonstrated that MISL results in lower autocorrelation sidelobes with less computational complexity.

The good correlation property of a single unimodular sequence is also extended to MIMO systems, where multiple sequences are transmitted. The good autocorrelation is defined for each sequence as that for a single sequence. Meanwhile, good cross-correlation demands that any sequence be nearly uncorrelated with time-shifted versions of the other sequences. In [3], algorithms CA-direct (CAD) and CA-new (CAN) are developed to obtain sequence sets of low auto- and cross-correlation sidelobes. Also [7] proposes some efficient algorithms to minimize the same metric in [3].

The aforementioned ISL and ISL-related metrics are both alternative ways to describe the impulse-like correlation characteristics. Sequences with such properties enable matched filters at the receiver side to easily extract the signals backscattered from the range bin of interest and attenuate signals backscattered from other range bins [3]. Nevertheless, matched filters take no advantage of any prior information on the channel when the unimodular low-ISL sequences are used for estimation.

The unimodular constraint is actually a special case of the low PAR constraint, which imposes how the largest amplitude of the sequence compares with its average power. The low PAR constraint, as a structural requirement, has been well studied in the design of tight frames [2]. Although the individual vector norms of a frame could be adjusted to maximize the sum-capacity of DS-CDMA links, the optimality in terms of any performance measures was not directly considered therein. Furthermore, the algorithm they proposed is based on alternating projection that often suffers a slow convergence.

As far as channel estimation is concerned, many studies have been conducted for both frequency-flat and frequency-selective fading channels under minimum mean square error (MMSE) estimation and conditional mutual information (CMI) maximization. Most of those obtained optimization problems, however, only address the power constraint without addressing the unimodular or low PAR constraints. In [8]–[12], training sequence design for flat MIMO channels is studied assuming some special structures, e.g., Kronecker product, on the prior covariance matrices of channel and noise. It is shown that the optimization problem can be reformulated as power allocations

This work was supported by the Hong Kong RGC 16206315 research grant. Zhongju Wang and Daniel P. Palomar are with the Hong Kong University of Science and Technology (HKUST), Hong Kong. E-mail: {zwangaq, palomar}@ust.hk. Prabhu Babu was with the Hong Kong University of Science and Technology (HKUST), Hong Kong. He is now with CARE, IIT Delhi, Delhi, India. Email: prabhubabu@care.iitd.ac.in.

using the majorization theory, and the waterfilling solutions are obtained. Meanwhile, problems of similar formulations have also been studied in joint linear transmitter-receiver design [13], [14]. To deal with arbitrarily correlated MIMO channels, some numerical algorithms based on block coordinate descent are proposed in [15]–[17].

More related to our work is training sequence design for frequency-selective fading channels. Under a total power constraint, channel capacity is investigated for SISO channels [18] and MIMO channels [19]. Independent and identically distributed channel coefficients and noise are assumed to facilitate mathematical analysis. As a result, impulse-like sequences for both types of channel are suggested for optimal estimation. Optimal design for the MMSE channel estimation has been studied in [20], where the noise is assumed to be white and the channel taps are uncorrelated; however, such assumption is hardly satisfied in practice. And there is no guarantee of finding an optimal solution for an arbitrary length of training or channel correlation. More important, their results cannot be used when the unimodular or low PAR constraint is imposed on the sequences to be designed.

We formulate the problem as the design of optimal unimodular sequences based on the MMSE and the CMI. Both problems are non-convex with the bothersome unimodular constraint. Without assuming any amenable structures, e.g., Kronecker product, on the prior channel and noise covariances, the problems are also challenging even if only the power constraint is imposed. To tackle those issues, the majorization-minimization (MM) technique is employed to develop efficient algorithms. By rewriting the objective functions in a more appropriate way, majorizing/minorizing functions can be obtained for minimization/maximization objective. As a result, the original problems are solved instead by a sequence of simple problems, each of which turns out to have a closed-form solution. Convergence of our proposed algorithms is guaranteed, and an acceleration scheme is also given to improve the convergence rate. For low PAR constraints, similar problems can be formulated, and the developed algorithms need only a few modifications to be applied.

The rest of this paper is organized as follows. In Section II, the channel model is described, on which the optimal unimodular sequence design problems are formulated. In Section III, derivations of algorithms for both the MMSE minimization and the CMI maximization are presented, followed by a brief analysis of convergence properties and an acceleration scheme. The optimal design under the low PAR constraints is discussed in IV. Numerical examples are presented in Section V. And conclusion is then given in Section VI.

Notation: Scalars are represented by italic letters. Boldface uppercase and lowercase letters denote matrices and vectors, respectively. \mathbb{C} is the set of complex numbers. The identity matrix is denoted by \mathbf{I} with the size implicit in the context if undeclared. The superscripts $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ denote respectively transpose, conjugate transpose and complex conjugate. With $\text{vec}(\mathbf{X})$, the vector is formed by stacking the columns of \mathbf{X} . The Kronecker product is denoted by \otimes . $\mathbb{E}(\cdot)$ takes the expectation of random variable. $\text{Tr}(\cdot)$ is the trace of a matrix. $\|\cdot\|_F$ is Frobenius norm of a matrix.

II. CHANNEL MODEL AND PROBLEM FORMULATIONS

We consider a block-fading or quasistatic multi-input multi-output (MIMO) channel. Assume the number of transmit antennas and receive antennas are N_t and N_r , respectively, and the channel impulse response is described as a length- $(K+1)$ sequence of matrices $\mathbf{H}_0, \dots, \mathbf{H}_K \in \mathbb{C}^{N_r \times N_t}$. In the training period, a length- N sequence is sent through the channel from each transmit antenna or, equivalently, a length- N_t vector \mathbf{u}_n from the set of transmit antennas at the time instant $n = 1, \dots, N$. For simplicity, we still call this sequence of vectors as a sequence, which is denoted by $\mathbf{U} = [\mathbf{u}_{n,m}] = [\mathbf{u}_1 \ \dots \ \mathbf{u}_N]^T \in \mathbb{C}^{N \times N_t}$. Considering the unimodular constraint with energy budget $\text{Tr}(\mathbf{U}^H \mathbf{U}) = \alpha$, we want to design $\mathbf{U} \in \mathcal{U}$, where

$$\mathcal{U} = \left\{ \mathbf{U} \middle| |u_{n,m}| = \sqrt{\frac{\alpha}{NN_t}}, n = 1, \dots, N; m = 1, \dots, N_t \right\}. \quad (1)$$

And the received sequence is given by

$$\mathbf{y}_n = \sum_{k=0}^K \mathbf{H}_k \mathbf{u}_{n-k} + \mathbf{v}_n, \quad (2)$$

where $\mathbf{u}_n = \mathbf{0}$ when $n \leq 0$ or $n > N$, and \mathbf{v}_n is an $N_r \times 1$ noise vector. Equation (2) can be written in a matrix form as

$$\begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_{N+K}^T \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T & \mathbf{0} \\ \vdots & \ddots & \mathbf{u}_1^T \\ \vdots & \ddots & \vdots \\ \mathbf{u}_N^T & \ddots & \vdots \\ \mathbf{0} & \ddots & \mathbf{u}_N^T \end{bmatrix} \begin{bmatrix} \mathbf{H}_0^T \\ \vdots \\ \mathbf{H}_K^T \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_{N+K}^T \end{bmatrix}. \quad (3)$$

Let $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_{N+K}]^T \in \mathbb{C}^{(N+K) \times N_r}$ be the received matrix, and

$$\mathbf{S} = \mathcal{T}(\mathbf{U}) = \begin{bmatrix} \mathbf{u}_1^T & \mathbf{0} \\ \vdots & \ddots & \mathbf{u}_1^T \\ \vdots & \ddots & \vdots \\ \mathbf{u}_N^T & \ddots & \vdots \\ \mathbf{0} & \ddots & \mathbf{u}_N^T \end{bmatrix} \in \mathbb{C}^{(N+K) \times (K+1)N_t} \quad (4)$$

be a block Toeplitz convolution matrix with $[\mathbf{U}^T \ \mathbf{0}]^T$ being the first block and remaining blocks are obtained by a downward circular shift of the previous block. Note that since $\text{Tr}(\mathbf{U}^H \mathbf{U}) = \alpha$, then $\text{Tr}(\mathbf{S}^H \mathbf{S}) = \alpha(K+1)$. $\mathbf{H} = [\mathbf{H}_0 \ \dots \ \mathbf{H}_K]^T \in \mathbb{C}^{(K+1)N_t \times N_r}$ is the channel impulse response with matrix-form taps, and \mathbf{V} is the noise matrix. Thus, we can write in a compact way the received signal as

$$\mathbf{Y} = \mathbf{S}\mathbf{H} + \mathbf{V}. \quad (5)$$

It can be easily seen that each column of \mathbf{Y} corresponds to a received sequence for one of the N_r receive antennas, i.e., a multi-input single-output (MISO) channel. Let $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{h} = \text{vec}(\mathbf{H})$, and $\mathbf{v} = \text{vec}(\mathbf{V})$, and based on $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X}) \text{vec}(\mathbf{Y})$, we have

$$\mathbf{y} = (\mathbf{I}_{N_r} \otimes \mathbf{S}) \mathbf{h} + \mathbf{v}. \quad (6)$$

A. Heuristic Existing Methods

Most of the current works on unimodular sequence design focus on good auto- and cross-correlation properties; see [3] on MIMO radar unimodular codes and references therein. The good correlation properties are particularly desired in that the matched filter is employed in subsequent channel estimation. As a matter of fact, the obtained channel estimate is closely related to maximum likelihood (ML) estimation. Assume the vectorized noise in the channel model (6) follows a circularly complex Gaussian distribution, $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. Minimizing the mean square error (MSE) $E\{\|\hat{\mathbf{h}}_{\text{ML}} - \mathbf{h}\|^2\}$ results in the ML channel estimate [21]

$$\hat{\mathbf{h}}_{\text{ML}} = \left((\mathbf{I}_{N_r} \otimes \mathbf{S})^H (\mathbf{I}_{N_r} \otimes \mathbf{S}) \right)^{-1} (\mathbf{I}_{N_r} \otimes \mathbf{S})^H \mathbf{y} \quad (7)$$

$$= (\mathbf{I}_{N_r} \otimes \mathbf{S}^H \mathbf{S})^{-1} (\mathbf{I}_{N_r} \otimes \mathbf{S})^H \mathbf{y}, \quad (8)$$

where the second equality is due to $(\mathbf{X} \otimes \mathbf{Y})(\mathbf{M} \otimes \mathbf{N}) = \mathbf{X}\mathbf{M} \otimes \mathbf{Y}\mathbf{N}$. And the corresponding error is given by

$$\mathcal{E} = \text{Tr} \left(\left((\mathbf{I}_{N_r} \otimes \mathbf{S})^H (\mathbf{I}_{N_r} \otimes \mathbf{S}) \right)^{-1} \right) \quad (9)$$

$$= \text{Tr} \left((\mathbf{I}_{N_r} \otimes \mathbf{S}^H \mathbf{S})^{-1} \right) \quad (10)$$

$$= N_r \text{Tr} \left((\mathbf{S}^H \mathbf{S})^{-1} \right). \quad (11)$$

To minimize the error of ML estimation, the training sequence should be a solution to the optimization problem

$$\underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} \quad \mathcal{E} \quad \text{subject to} \quad \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}. \quad (12)$$

Lemma 1 [22]. *Let $\mathbf{X} \in \mathbb{C}^{M \times N}$ be such that $\text{Tr}(\mathbf{X}^H \mathbf{X}) \leq \mu$ for some constant μ . The minimum of $\text{Tr}((\mathbf{X}^H \mathbf{X})^{-1})$ is achieved when $\mathbf{X}^H \mathbf{X} = \frac{\mu}{N} \mathbf{I}$, provided that inverse of $\mathbf{X}^H \mathbf{X}$ exists.*

An approximation to problem (12) is as follows. According to Lemma 1, the objective function (11) is minimized when $(\text{Tr}(\mathbf{S}^H \mathbf{S}) = \alpha(K+1))$

$$\mathbf{S}^H \mathbf{S} = \frac{\alpha}{N_t} \mathbf{I}_{(K+1)N_t}, \quad (13)$$

if only the energy constraint is considered. Therefore, a heuristic approximation of the ML optimal sequence design could be formulated as

$$\underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} \quad \left\| \mathbf{S}^H \mathbf{S} - \frac{\alpha}{N_t} \mathbf{I} \right\|_F^2 \quad (14)$$

subject to $\mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}.$

The optimal \mathbf{S} satisfying (13) portrays an impulse-like correlation shape pursued in [3], [7], where the aperiodic cross-correlation is defined as

$$r_{m_1, m_2}(k) = \sum_{n=k+1}^N u_{n, m_1} u_{n-k, m_2}^* \quad (15)$$

for $m_1, m_2 = 1, \dots, N_t$ and lags $k = 0, \dots, N-1$. Equation (15) also defines the autocorrelation for the sequence of each transmit antenna when $m_1 = m_2$. Accordingly, the correlation

matrices for different lags $k = -(N-1), \dots, 0, \dots, (N-1)$ are given by

$$\Sigma_k = \begin{bmatrix} r_{1,1}(k) & r_{1,2}(k) & \cdots & r_{1,N_t}(k) \\ r_{2,1}(k) & r_{2,2}(k) & \cdots & r_{2,N_t}(k) \\ \vdots & \vdots & \ddots & \vdots \\ r_{N_t,1}(k) & r_{N_t,2}(k) & \cdots & r_{N_t,N_t}(k) \end{bmatrix}, \quad (16)$$

with $r_{m_1, m_2}(-k) = r_{m_1, m_2}^*(k)$, and $\Sigma_{-k} = \Sigma_k^H$. Let us define the correlation matrix for a sequence \mathbf{S} as

$$\Sigma = \mathbf{S}^H \mathbf{S}, \quad (17)$$

and then we have

$$\Sigma = \begin{bmatrix} \Sigma_0 & \Sigma_{-1} & \cdots & \Sigma_{-K} \\ \Sigma_1 & \Sigma_0 & \cdots & \Sigma_{-(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_K & \Sigma_{K-1} & \cdots & \Sigma_0 \end{bmatrix}. \quad (18)$$

Note that Σ only describes correlations at lags of interest, which in this case is determined by the length of the channel impulse response. To achieve the optimality dictated by (13), we can rewrite approximation problem (14) as

$$\underset{\mathbf{U}}{\text{minimize}} \quad (K+1) \left\| \Sigma_0 - \frac{\alpha}{N_t} \mathbf{I} \right\|_F^2 + 2 \sum_{k=1}^K (K+1-k) \|\Sigma_k\|_F^2 \quad (19)$$

subject to $\mathbf{U} \in \mathcal{U}.$

The objective function of (19) is indeed the weighted correlation minimization criterion within the lag interval $k = 0, \dots, K$ [3], for which algorithms WeCan and CAD were proposed. Another formulation is also presented in a similar attempt to procure the good correlation property as

$$\underset{\mathbf{U}}{\text{minimize}} \quad \left\| \Sigma_0 - \frac{\alpha}{N_t} \mathbf{I} \right\|_F^2 + 2 \sum_{k=1}^{N-1} \|\Sigma_k\|_F^2 \quad (20)$$

subject to $\mathbf{U} \in \mathcal{U},$

for which an algorithm called CAN was developed in [5], and WeCAN can be employed as well. In [7], both problems (19) and (20) were studied by considering a more general weighted formulation, and efficient algorithms were proposed.

We can see that sequences with good auto- and cross-correlation properties are desirable in general as no prior information on the channel is taken into account in the ensuing channel estimation task. Channel statistics, however, are often available on both the transmitter sides and receiver sides, and incorporating those priors into the design of the training sequence will improve the performance of channel estimator. In the following subsections, we will formulate the unimodular sequence design problem based on the MMSE minimization and the CMI maximization, both of which have been adopted as criteria in various estimation problems. In order for the channel model (6) to be general, we assume $\mathbf{h} \sim \mathcal{CN}(\mathbf{h}_0, \mathbf{R}_0)$, and the noise $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{W})$, where both the channel covariance \mathbf{R}_0 and the noise covariance \mathbf{W} are arbitrary.

B. Optimal Sequence Design by Minimizing the MMSE Criterion

Given the channel model (6), by minimizing the MSE $E\{\|\hat{\mathbf{h}}_{\text{MMSE}} - \mathbf{h}\|^2\}$, the MMSE estimator of the channel impulse \mathbf{h} is given by

$$\hat{\mathbf{h}}_{\text{MMSE}} = \mathbf{R}_0 \tilde{\mathbf{S}}^H \left(\tilde{\mathbf{S}} \mathbf{R}_0 \tilde{\mathbf{S}}^H + \mathbf{W} \right)^{-1} \left(\mathbf{y} - \tilde{\mathbf{S}} \mathbf{h}_0 \right) + \mathbf{h}_0, \quad (21)$$

where $\tilde{\mathbf{S}} = \mathbf{I}_{N_r} \otimes \mathbf{S}$ [21]. And the error covariance matrix is

$$\mathbf{R} = E \left\{ \left(\hat{\mathbf{h}}_{\text{MMSE}} - \mathbf{h} \right) \left(\hat{\mathbf{h}}_{\text{MMSE}} - \mathbf{h} \right)^H \right\} \quad (22)$$

$$= \mathbf{R}_0 - \mathbf{R}_0 \tilde{\mathbf{S}}^H \left(\tilde{\mathbf{S}} \mathbf{R}_0 \tilde{\mathbf{S}}^H + \mathbf{W} \right)^{-1} \tilde{\mathbf{S}} \mathbf{R}_0 \quad (23)$$

$$= \left(\mathbf{R}_0^{-1} + \tilde{\mathbf{S}}^H \mathbf{W}^{-1} \tilde{\mathbf{S}} \right)^{-1}, \quad (24)$$

where the last equality is due to the matrix inversion lemma [13]. The MMSE is thus given by

$$\text{MMSE}(\mathbf{S}) = \text{Tr}(\mathbf{R}), \quad (25)$$

and the following problem can be formulated

$$\underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} \quad \text{MMSE}(\mathbf{S}) \quad \text{subject to} \quad \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}, \quad (26)$$

which gives the optimal unimodular training sequence for the MMSE channel estimation.

C. Optimal Sequence Design by Maximizing the CMI Criterion

Apart from the MMSE criterion, another popular statistical measure in channel estimation is the conditional mutual information (CMI) between the channel impulse response and the received sequence, e.g., [8]. The CMI is defined as

$$\text{CMI}(\mathbf{S}) = I(\mathbf{h}; \mathbf{y} | \mathbf{S}) \quad (27)$$

$$= H(\mathbf{h}) - H(\mathbf{h} | \mathbf{y}, \mathbf{S}), \quad (28)$$

where $H(\cdot)$ is the differential entropy of a distribution [23]. Under the linear model (6) with Gaussian distributed channel impulse and noise, we have the conditional distribution $\mathbf{h} | \mathbf{y}, \mathbf{S} \sim \mathcal{CN}(\hat{\mathbf{h}}, \mathbf{R})$. Then $\text{CMI}(\mathbf{S})$ can be written as

$$\begin{aligned} \text{CMI}(\mathbf{S}) &= \frac{1}{2} \log \left((2\pi e)^{(K+1)N_t N_r} \det(\mathbf{R}_0) \right) \\ &\quad - \frac{1}{2} \log \left((2\pi e)^{(K+1)N_t N_r} \det(\mathbf{R}) \right) \end{aligned} \quad (29)$$

$$= \frac{1}{2} \log \det(\mathbf{R}_0 \mathbf{R}^{-1}). \quad (30)$$

By maximizing $\text{CMI}(\mathbf{S})$ we reach the following optimization problem

$$\underset{\mathbf{U}, \mathbf{S}}{\text{maximize}} \quad \text{CMI}(\mathbf{S}) \quad \text{subject to} \quad \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}. \quad (31)$$

Remark: It should be mentioned that channel model (6) includes the SISO channel as a special case. A lot of efforts have been made to construct unimodular sequences for SISO channels via either analytical methods or computational approaches. Apart from early works on binary sequences and polyphase sequences, e.g., [24], [25], numerical algorithms are

provided to design unimodular sequences of good correlation properties [5], [6]. Let $N_t = N_r = 1$ and \mathbf{u} denote the training sequence, then $\mathbf{S} = \mathcal{T}(\mathbf{u})$ is a Toeplitz convolution matrix and expression (16) reduces to a scalar that gives autocorrelations at different lags for the sequence \mathbf{u} . And similar formulations as (19) and (20) are proposed in order to obtain sequences of good autocorrelation properties. However, as we have seen in the previous discussion, the resulting channel estimate cannot benefit from the available knowledge of channel statistics. Therefore, designing optimal training sequences by minimizing the MMSE criterion or maximizing the CMI criterion will be beneficial in terms of final estimation performances. Without any modifications, formulations (26) and (31) can be deployed in the context of SISO channels.

III. ALGORITHMS FOR UNIMODULAR SEQUENCE DESIGN

In this section, we develop efficient algorithms to solve problems (26) and (31). There is an extensive literature dealing with optimization problems of similar objective functions with only power constraint on \mathbf{S} where, assuming some special structure for the prior covariance matrices of channel and noise, the problems are reformulated as power allocation with waterfilling-like solutions. In our formulations, however, it is not only the Toeplitz structure of \mathbf{S} but also the tough unimodular constraint that prevents us from adopting the same approach.

It is worth mentioning that a possible approach to problem (26) is a two-stage procedure [26] related to correlation shaping. If the channel noise is independent and identically distributed, i.e., $\mathbf{W} = \sigma^2 \mathbf{I}$ for some power density σ^2 , the objective function becomes

$$\text{MMSE}(\mathbf{S}) = \text{Tr} \left(\left(\mathbf{R}_0^{-1} + \frac{1}{\sigma^2} \tilde{\mathbf{S}}^H \tilde{\mathbf{S}} \right)^{-1} \right) \quad (32)$$

$$= \text{Tr} \left(\left(\mathbf{R}_0^{-1} + \frac{1}{\sigma^2} \mathbf{I}_{N_r} \otimes \mathbf{\Sigma} \right)^{-1} \right), \quad (33)$$

where the second equality follows from substitution of the correlation matrix $\mathbf{\Sigma} = \mathbf{S}^H \mathbf{S}$. Consider only the constraint $\text{Tr}(\mathbf{\Sigma}) = (K+1)\alpha$ induced by the energy budget in (26), minimizing (33) with respect to $\mathbf{\Sigma}$ (instead of \mathbf{S}) can be rewritten as an SDP by resorting to the Schur-complement theorem [27], which yields the optimal correlation matrix $\mathbf{\Sigma}^*$. Once $\mathbf{\Sigma}^*$ is obtained, the problem boils down to recovering sequences from its correlation matrix, which is to solve the following approximation problem

$$\begin{aligned} &\underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} \quad \|\mathbf{S}^H \mathbf{S} - \mathbf{\Sigma}^*\|_F \\ &\text{subject to} \quad \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}, \end{aligned} \quad (34)$$

if zero error is achievable. For a single sequence and without the unimodular constraint, problem (34) can be tackled by means of filter design [28]. However, constructing a unimodular sequence that presents a prescribed correlation shape is challenging. As a special case, [3] and [7] have studied this problem only when the correlation matrix is an identity. On the other hand, it is not guaranteed that the objective in (34) can reach zero when minimized. For example, when the number

of sequences is relatively large for the training length, it is impossible to design sequences such that correlation matrix is an identity, i.e., auto- and cross-correlation cannot be made small simultaneously [3].

Therefore, it is advisable to solve problems (26) and (31) directly with the colored noise considered. In the following, we will devise algorithms for both problems based on the majorization-minimization framework.

A. Majorization-Minimization Framework

The majorization-minimization, or MM method is a general framework for solving an optimization problem indirectly. In this section, we will briefly introduce the idea of the MM method for a minimization problem, and the details can be found in [29], [30].

The MM method tackles a difficult optimization problem by solving a series of simple approximation problems. Given a minimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in \mathcal{X}, \quad (35)$$

and a feasible starting point $\mathbf{x}^{(0)} \in \mathcal{X}$, the MM method minimizes a sequence of surrogate functions $g(\mathbf{x}, \mathbf{x}^{(t)})$, $t = 0, 1, \dots$ instead. Each surrogate function is a majorization function of $f(\mathbf{x})$ at $\mathbf{x}^{(t)}$ that satisfies:

$$g(\mathbf{x}^{(t)}, \mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \quad (36)$$

$$g(\mathbf{x}, \mathbf{x}^{(t)}) \geq f(\mathbf{x}) \quad \text{for every } \mathbf{x} \in \mathcal{X}, \quad (37)$$

and

$$\mathbf{x}^{(t+1)} \in \underset{\mathbf{x} \in \mathcal{X}}{\arg \min} \quad g(\mathbf{x}, \mathbf{x}^{(t)}). \quad (38)$$

According to the rules (36) and (37), we have

$$f(\mathbf{x}^{(t+1)}) \leq g(\mathbf{x}^{(t+1)}, \mathbf{x}^{(t)}) \leq g(\mathbf{x}^{(t)}, \mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \quad (39)$$

and consequently, the MM method produces a sequence of points $\mathbf{x}^{(t)}$, for which the original objective function of (35) is monotonically decreased. Provided that the objective function is bounded below, it is guaranteed that the MM algorithm will converge to a stationary point.

The key question is then how to find a good majorization function $g(\mathbf{x}, \mathbf{x}^{(t)})$ such that the resulting problems (38) are easy to solve. Although there is no universal rule to determine the function $g(\mathbf{x}, \mathbf{x}^{(t)})$, the structure of the problem at hand can nevertheless provide helpful hints and some tricks are suggested in [29].

B. MM-Based Algorithms

Let us introduce $\mathbf{P} = \tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H + \mathbf{W}$, and by (23) the objective function for the MMSE minimization problem (26) can be written as

$$\text{MMSE}(\mathbf{S}) = \text{Tr}(\mathbf{R}_0 - \mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{P}^{-1}\tilde{\mathbf{S}}\mathbf{R}_0). \quad (40)$$

Lemma 2. *The function $f(\mathbf{X}, \mathbf{Z}) = \text{Tr}(\mathbf{X}^H\mathbf{Z}^{-1}\mathbf{X})$ is a matrix fractional function and is jointly convex in $\mathbf{Z} \succ 0$ and \mathbf{X} [27].*

By Lemma 2, $\text{MMSE}(\mathbf{S}) = \text{Tr}(\mathbf{R}_0 - \mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{P}^{-1}\tilde{\mathbf{S}}\mathbf{R}_0)$ is jointly concave in $\{\tilde{\mathbf{S}}, \mathbf{P}\}$ (recall that $\tilde{\mathbf{S}} = \mathbf{I}_{N_r} \otimes \mathbf{S}$). Since a concave function is upper-bounded by its supporting hyperplane, $\text{MMSE}(\mathbf{S})$ can be majorized as follows:

$$\text{MMSE}(\mathbf{S}) \leq g_{\text{MMSE}}(\mathbf{S}, \mathbf{S}^{(t)}) \quad (41)$$

$$= \text{MMSE}(\mathbf{S}^{(t)}) + \text{Tr}\left(\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H \mathbf{A}^{(t)}\right) - 2\text{Re}\left\{\text{Tr}\left(\mathbf{R}_0\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}\right)\right\}, \quad (42)$$

where $\tilde{\mathbf{S}}^{(t)} = \mathbf{I}_{N_r} \otimes \mathbf{S}^{(t)}$ with $\mathbf{S}^{(t)} = \mathcal{T}(\mathbf{U}^{(t)})$, and $\mathbf{A}^{(t)} = \left(\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\left(\tilde{\mathbf{S}}^{(t)}\right)^H + \mathbf{W}\right)^{-1}\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0$. To solve problem (26), it suffices to solve iteratively the following problem:

$$\begin{aligned} &\underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} \quad g_{\text{MMSE}}(\mathbf{S}, \mathbf{S}^{(t)}) \\ &\text{subject to} \quad \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}, \end{aligned} \quad (43)$$

For problem (31), the objective function can be written as

$$\text{CMI}(\mathbf{S}) = \frac{1}{2} \log \det \left(\mathbf{R}_0 \left(\mathbf{R}_0 - \mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{P}^{-1}\tilde{\mathbf{S}}\mathbf{R}_0 \right)^{-1} \right). \quad (44)$$

Lemma 3. *Given a positive semidefinite matrix \mathbf{M} , the function $h(\mathbf{Z}, \mathbf{X}) = \mathbf{M} - \mathbf{M}\mathbf{X}^H\mathbf{Z}^{-1}\mathbf{X}\mathbf{M}$ is matrix concave over \mathbf{X} of an appropriate size and $\mathbf{Z} \succ 0$ [27]. Since $-\log \det(\cdot)$ is matrix convex and decreasing over positive definite cone, $-\log \det(\mathbf{M} - \mathbf{M}\mathbf{X}^H\mathbf{Z}^{-1}\mathbf{X}\mathbf{M})$ is convex in $\{\mathbf{Z}, \mathbf{X}\}$.*

Owing to Lemma 3, $\text{CMI}(\mathbf{S})$ is jointly convex in $\{\tilde{\mathbf{S}}, \mathbf{P}\}$, and we can obtain the following minorization

$$\text{CMI}(\mathbf{S}) \geq g_{\text{CMI}}(\mathbf{S}, \mathbf{S}^{(t)}) \quad (45)$$

$$= \text{Re}\left\{\text{Tr}\left(\mathbf{R}_0\left(\mathbf{R}^{(t)}\right)^{-1}\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}\right)\right\} \quad (46)$$

$$= -\frac{1}{2}\text{Tr}\left(\left(\mathbf{R}^{(t)}\right)^{-1}\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{A}^{(t)}\right) + \text{CMI}(\mathbf{S}^{(t)}), \quad (47)$$

where

$$\mathbf{R}^{(t)} = \mathbf{R}_0 - \mathbf{R}_0\left(\tilde{\mathbf{S}}^{(t)}\right)^H \left(\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\left(\tilde{\mathbf{S}}^{(t)}\right)^H + \mathbf{W}\right)^{-1} \tilde{\mathbf{S}}^{(t)}\mathbf{R}_0 \quad (48)$$

$$= \mathbf{R}_0^{-1} + \left(\tilde{\mathbf{S}}^{(t)}\right)^H \mathbf{W}^{-1}\tilde{\mathbf{S}}^{(t)}. \quad (49)$$

As a result, solving the CMI maximization problem (31) is equivalent to solving the series of minorized problems

$$\begin{aligned} &\underset{\mathbf{U}, \mathbf{S}}{\text{maximize}} \quad g_{\text{CMI}}(\mathbf{S}, \mathbf{S}^{(t)}) \\ &\text{subject to} \quad \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}. \end{aligned} \quad (50)$$

Notice that problems (43) and (50) share a similar form of objective function. Let

$$\begin{aligned} g(\mathbf{S}; \mathbf{S}^{(t)}, \mathbf{V}^{(t)}) &= \text{Tr}\left(\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{A}^{(t)}\right) \\ &\quad - 2\text{Re}\left\{\text{Tr}\left(\mathbf{R}_0\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}\right)\right\}, \end{aligned} \quad (51)$$

where $\mathbf{V}^{(t)} = \mathbf{I}$ for the MMSE minimization problem and $\mathbf{V}^{(t)} = (\mathbf{R}^{(t)})^{-1}$ for the CMI maximization problem. After reversing the sign of objective function of (50) and ignoring the constants and the scaling factor, the following unified problem is obtained

$$\begin{aligned} & \underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} && g(\mathbf{S}; \mathbf{S}^{(t)}, \mathbf{V}^{(t)}) \\ & \text{subject to} && \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}. \end{aligned} \quad (52)$$

Lemma 4. Given Hermitian $\mathbf{M} \in \mathbb{C}^{n \times n}$ and $\mathbf{Z} \in \mathbb{C}^{m \times m}$ and any $\mathbf{X}^{(t)} \in \mathbb{C}^{m \times n}$, the function $\text{Tr}(\mathbf{Z}\mathbf{X}\mathbf{M}\mathbf{X}^H)$ can be majorized by $-2\text{Re}\left\{\text{Tr}\left((\lambda\mathbf{X}^{(t)} - \mathbf{Z}\mathbf{X}^{(t)}\mathbf{M})^H \mathbf{X}\right)\right\} + \lambda\|\mathbf{X}\|_F^2 + \text{const}$, where $\lambda\mathbf{I} \succeq \mathbf{M}^T \otimes \mathbf{Z}$ for some constant λ .

Proof: Given $\lambda\mathbf{I} \succeq \mathbf{M}^T \otimes \mathbf{L}\mathbf{L}^H$ for some constant λ , we have

$$\begin{aligned} & \text{Tr}(\mathbf{Z}\mathbf{X}\mathbf{M}\mathbf{X}^H) \\ &= \text{vec}^H(\mathbf{X}) \text{vec}(\mathbf{Z}\mathbf{X}\mathbf{M}) \end{aligned} \quad (53)$$

$$= \text{vec}^H(\mathbf{X}) (\mathbf{M}^T \otimes \mathbf{Z}) \text{vec}(\mathbf{X}) \quad (54)$$

$$\begin{aligned} & \leq -2\text{Re}\left\{\text{vec}^H(\mathbf{X}) (\lambda\mathbf{I} - \mathbf{M}^T \otimes \mathbf{Z}) \text{vec}(\mathbf{X}^{(t)})\right\} \\ & \quad + \text{vec}^H(\mathbf{X}^{(t)}) (\lambda\mathbf{I} - \mathbf{M}^T \otimes \mathbf{Z}) \text{vec}(\mathbf{X}^{(t)}) \\ & \quad + \lambda \text{vec}^H(\mathbf{X}) \text{vec}(\mathbf{X}) \end{aligned} \quad (55)$$

$$\begin{aligned} &= -2\text{Re}\left\{\text{Tr}\left(\lambda\mathbf{X}^H\mathbf{X}^{(t)} - \mathbf{Z}\mathbf{X}^{(t)}\mathbf{M}\mathbf{X}^H\right)\right\} \\ & \quad + \lambda\|\mathbf{X}\|_F^2 + \text{vec}^H(\mathbf{X}^{(t)}) (\lambda\mathbf{I} - \mathbf{M}^T \otimes \mathbf{Z}) \text{vec}(\mathbf{X}^{(t)}). \end{aligned} \quad (56)$$

Notice that the third term of the last equation is simply a constant. And a scalar version of Lemma 4 can be found in [7, Lemma 1]. ■

To solve problem (52), yet a second majorization can be applied with Lemma 4 (note that $\|\tilde{\mathbf{S}}\|^2 = N_r(K+1)\alpha$):

$$\begin{aligned} & g(\mathbf{S}; \mathbf{S}^{(t)}, \mathbf{V}^{(t)}) \\ & \leq -2\text{Re}\left\{\text{Tr}\left(\lambda^{(t)}\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}^{(t)} - \mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\right.\right. \\ & \quad \left.\left.\tilde{\mathbf{S}}^H\right)\right\} - 2\text{Re}\left\{\text{Tr}\left(\mathbf{R}_0\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\tilde{\mathbf{S}}\right)\right\} + \text{const} \quad (57) \\ &= -2\text{Re}\left\{\text{Tr}\left((\lambda^{(t)}\tilde{\mathbf{S}}^{(t)} - \mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\right.\right. \\ & \quad \left.\left.+ \mathbf{A}^{(t)}\mathbf{V}^{(t)}\mathbf{R}_0)^H\tilde{\mathbf{S}}\right)\right\} + \text{const}, \end{aligned} \quad (58)$$

where $\lambda^{(t)}\mathbf{I} \succeq \mathbf{R}_0^T \otimes \mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H$. The tightest upper bound will be $\lambda^{(t)} = \lambda_{\max}(\mathbf{R}_0^T \otimes \mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H)$. But computing the largest eigenvalue is costly especially when the size of the matrix is large, and thus an alternative is advisable. Since both \mathbf{R}_0 and $\mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H$ are positive semidefinite matrices, the largest eigenvalues are bounded as

$$\lambda_{\max}(\mathbf{R}_0) \leq \|\mathbf{R}_0\|_1, \quad (59)$$

$$\lambda_{\max}(\mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H) \leq \left\|\mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\right\|_1, \quad (60)$$

where $\|\cdot\|_1$ is maximum column sum matrix norm [31]. With $\lambda_{\max}(\mathbf{X} \otimes \mathbf{Z}) = \lambda_{\max}(\mathbf{X})\lambda_{\max}(\mathbf{Z})$, we propose

$$\lambda^{(t)} = \|\mathbf{R}_0\|_1 \left\|\mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\right\|_1. \quad (61)$$

Let $\mathbf{B}(\tilde{\mathbf{S}}^{(t)}, \mathbf{V}^{(t)}) = \lambda^{(t)}\tilde{\mathbf{S}}^{(t)} - \mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0 + \mathbf{A}^{(t)}\mathbf{V}^{(t)}\mathbf{R}_0$, and considering $\tilde{\mathbf{S}} = \mathbf{I}_{N_r} \otimes \mathbf{S}$ with $\mathbf{S} = \mathcal{T}(\mathbf{U})$, we have

$$\mathbf{U}^{(t+1)} \in \arg \min_{|u_{n,m}|=\sqrt{\frac{\alpha}{NN_t}}} -2\text{Re}\left\{\text{Tr}\left(\left(\sum_{i,j}\mathbf{B}[i,j]\right)^H\mathbf{U}\right)\right\}, \quad (62)$$

where $\mathbf{B}[i,j]$ is a submatrix of \mathbf{B} with rows from $(N+K)(i-1)+j$ to $(N+K)(i-1)+N+j-1$ and columns from $N_t(K+1)(i-1)+N_t(j-1)+1$ to $N_t(K+1)(i-1)+N_tj$, for $i = 1, \dots, N_r$, and $j = 1, \dots, K+1$. To find the next update $\mathbf{U}^{(t+1)}$, note that (62) can be equivalently written as

$$\mathbf{U}^{(t+1)} \in \arg \min_{|u_{n,m}|=\sqrt{\frac{\alpha}{NN_t}}} \left\|\mathbf{U} - \sum_{i,j}\mathbf{B}[i,j]\right\|_F^2. \quad (63)$$

And the minimum is achieved by projection onto a complex circle, which is

$$\mathbf{U}^{(t+1)} = \sqrt{\frac{\alpha}{NN_t}} e^{j \arg(\sum_{i,j}\mathbf{B}[i,j])}, \quad (64)$$

where $\arg(\cdot)$ is taken element-wise. The whole procedure is summarized in Algorithm 1. The iterations of the algorithm is deemed to be converged, e.g., when the difference between two consecutive updates for \mathbf{U} is no larger than some admitted threshold.

Algorithm 1 Design of unimodular training sequence for the MMSE minimization (26) or the CMI maximization (31).

-
- 1: Set $t = 0$, and initialize $u_{n,m}^{(0)}, n = 1, \dots, N; m = 1, \dots, N_t$.
 - 2: **repeat**
 - 3: $\mathbf{S}^{(t)} = \mathcal{T}(\mathbf{U}^{(t)})$, and $\tilde{\mathbf{S}}^{(t)} = \mathbf{I}_{N_r} \otimes \mathbf{S}^{(t)}$
 - 4: $\mathbf{A}^{(t)} = \left(\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0(\tilde{\mathbf{S}}^{(t)})^H + \mathbf{W}\right)^{-1}\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0$
 - 5: $\mathbf{V}^{(t)} = \begin{cases} \mathbf{I}, & \text{for the MMSE minimization} \\ \mathbf{R}^{(t)}, & \text{for the CMI maximization} \end{cases}$
 - 6: $\lambda^{(t)} = \|\mathbf{R}_0\|_1 \left\|\mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\right\|_1$
 - 7: $\mathbf{B}(\tilde{\mathbf{S}}^{(t)}, \mathbf{V}^{(t)}) = \lambda^{(t)}\tilde{\mathbf{S}}^{(t)} - \mathbf{A}^{(t)}\mathbf{V}^{(t)}(\mathbf{A}^{(t)})^H\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0 + \mathbf{A}^{(t)}\mathbf{V}^{(t)}\mathbf{R}_0$
 - 8: $\mathbf{U}^{(t+1)} = \sqrt{\frac{\alpha}{NN_t}} e^{j \arg(\sum_{i,j}\mathbf{B}[i,j])}$
 - 9: $t \leftarrow t + 1$
 - 10: **until** convergence
-

C. Convergence Analysis

Algorithm 1 is essentially based on the majorization-minimization framework, which has been shown to converge to a stationary point for bounded objective functions. The generated sequence of points $\mathbf{U}^{(t)}, t = 0, 1, \dots$, monotonically decreases or increases the objective function for minimization

and maximization problems, respectively. In this section, we give a detailed analysis of the convergence for Algorithm 1. Without loss of generality, we only consider minimizing the MMSE criterion.

For a constrained minimization problem with a smooth objective function, a stationary point is obtained when the following first-order optimality condition is satisfied.

Proposition 1. *Let $f : \mathbb{R}^N \rightarrow \mathbb{R}$ be a smooth function. A point \mathbf{x}^* is a local minimum of f within a subset $\mathcal{X} \subset \mathbb{R}^N$ if*

$$\nabla f(\mathbf{x}^*)^T \mathbf{y} \geq 0, \forall \mathbf{y} \in T_{\mathcal{X}}(\mathbf{x}^*), \quad (65)$$

where $T_{\mathcal{X}}(\mathbf{x}^*)$ is the tangent cone of \mathcal{X} at \mathbf{x}^* .

Provided Proposition 1, the convergence of our proposed algorithm is guaranteed as follows.

Theorem 1. *By solving the series of problems (62) in Algorithm 1, a sequence of points $\{\mathbf{U}^{(t)}, t = 0, \dots\}$ is obtained, of which every limit point is a stationary point of problem (26).*

Proof: A similar proof has been given in [6]. For details please refer to [6, Theorem 5]. ■

D. Accelerated Algorithm

To develop Algorithm 1 for solving problems (26) and (31), the original function was majorized/minorized twice, which may result in a loose surrogate function; see (41), (45) and (57). And the performance of the MM method is susceptible to the slow convergence as EM-like algorithms. Then following the same idea in [6], [7], we employ an off-the-shelf method, called squared iterative methods (SQUAREM) [32], to accelerate Algorithm 1. SQUAREM was originally proposed to improve the convergence of EM-type algorithms and simultaneously keep its simplicity and stability. It can be easily applied to accelerate the MM algorithms as well. For details of convergence analysis, also refer to [32]. Without loss of generality, we only consider acceleration of Algorithm 1 for the MMSE minimization problem. For the CMI maximization problem, a similar procedure can be followed.

Given the current point $\mathbf{U}^{(t)}$, we call iterative steps 3 to 8 of Algorithm 1 collectively as one MM update, denoted by $\text{MMupdate}(\mathbf{U}^{(t)})$. The accelerated computing scheme is given by Algorithm 2. The step length is chosen by the Cauchy-Barzilai-Borwein (CBB) method. And the back-tracking step is adopted to maintain the monotone property of generated iterates. To guarantee its feasibility, projection to the constrained set \mathcal{U} in steps 8 and 9 are applied.

IV. ALGORITHMS FOR SEQUENCE DESIGN UNDER PAR CONSTRAINTS

The unimodular constraint on the training sequence originates partly from the low peak-to-average power ratio (PAR) demand, e.g., in MIMO radar systems. Low PAR sequences have found many applications in practice because they can mitigate the non-linear effects at the transmitter side while enabling more flexibility of the designed sequences compared with unimodular ones. In this section, we consider the problem of designing optimal sequences with low PAR.

For a sequence of vectors $\mathbf{U} \in \mathbb{C}^{N \times N_t}$, $\mathbf{U}_{:,m}$ denotes the length- N sequence sent from the m th antenna, for $m =$

Algorithm 2 Accelerated scheme for designing optimal unimodular training sequence for the MMSE estimation.

```

1: Set  $t = 0$ , and initialize  $u_{n,m}^{(0)}, n = 1, \dots, N; m = 1, \dots, N_t$ .
2: repeat
3:    $\mathbf{U}_1 = \text{MMupdate}(\mathbf{U}^{(t)})$ 
4:    $\mathbf{U}_2 = \text{MMupdate}(\mathbf{U}_1)$ 
5:    $\mathbf{L}_1 = \mathbf{U}_1 - \mathbf{U}^{(t)}$ 
6:    $\mathbf{L}_2 = \mathbf{U}_2 - \mathbf{U}_1 - \frac{\|\mathbf{L}_1\|_F}{\|\mathbf{L}_2\|_F} \mathbf{L}_1$ 
7:   Step length  $l = -\frac{\|\mathbf{L}_1\|_F}{\|\mathbf{L}_2\|_F}$ 
8:    $\mathbf{U}^{(t+1)} = \sqrt{\frac{\alpha}{NN_t}} e^{j \arg(\mathbf{U}^{(t)} - 2l\mathbf{L}_1 + l^2\mathbf{L}_2)}$ 
9:   while  $\text{MMSE}(\mathbf{S}^{(t+1)}) > \text{MMSE}(\mathbf{S}^{(t)})$  do
10:      $l \leftarrow \frac{l-1}{2}$ , and go to step 8
11:    $t \leftarrow t + 1$ 
12: until convergence

```

$1, \dots, N_t$. And PAR is usually defined for each sequence transmitted by a single antenna as

$$\text{PAR}(\mathbf{U}_{:,m}) = \frac{\max_n \{|u_{n,m}|^2\}}{\frac{1}{N} \alpha_m}, \quad (66)$$

provided that the training energy for the m th antenna is $\|\mathbf{U}_{:,m}\|^2 = \alpha_m$. Determining training energy for each transmit antenna may depend on power distribution among antennas satisfying $\sum_{m=1}^{N_t} \alpha_m = \alpha$. And it follows that $1 \leq \text{PAR}(\mathbf{U}_{:,m}) \leq N$. When $\text{PAR}(\mathbf{U}_{:,m}) = 1$, PAR constraint reduces to the unimodular constraint. Given the PAR constraints for each transmit antenna

$$\text{PAR}(\mathbf{U}_{:,m}) \leq \xi_m, m = 1, \dots, N_t \quad (67)$$

the optimal sequence design problem for minimizing MMSE is then formulated as

$$\begin{aligned} & \underset{\mathbf{U}, \mathbf{S}}{\text{minimize}} && \text{MMSE}(\mathbf{S}) \\ & \text{subject to} && \mathbf{S} = \mathcal{T}(\mathbf{U}) \\ & && \|\mathbf{U}_{:,m}\|^2 = \alpha_m \\ & && \max_n \{|u_{n,m}|\} \leq \sqrt{\frac{\alpha_m \xi_m}{N}}, m = 1, \dots, N_t \end{aligned} \quad (68)$$

where $\text{MMSE}(\mathbf{S})$ is given by (40). For the CMI maximization, an optimization problem can be similarly formulated, which maximizes $\text{CMI}(\mathbf{S})$ (44) under the same constraints as that of (68).

Following the same procedure of applying the MM framework in Section III-B, the following majorized (minorized) problems can be obtained for problem (68) for the MMSE minimization (CMI maximization)

$$\begin{aligned} & \underset{\mathbf{U}}{\text{minimize}} && \left\| \mathbf{U} - \sum_{i,j} \mathbf{B}_{i,j} \right\|_F^2 \\ & \text{subject to} && \|\mathbf{U}_{:,m}\|^2 = \alpha_m \\ & && \max_n \{|u_{n,m}|\} \leq \sqrt{\frac{\alpha_m \xi_m}{N}}, m = 1, \dots, N_t \end{aligned} \quad (69)$$

It is obvious that problem (69) can be separated into N_t problems as

$$\begin{aligned} & \underset{\mathbf{U}_{:,m}}{\text{minimize}} \quad \|\mathbf{U}_{:,m} - \mathbf{c}_m\|^2 \\ & \text{subject to} \quad \|\mathbf{U}_{:,m}\|^2 = \alpha_m \\ & \quad \max_n \{|u_{n,m}|\} \leq \sqrt{\frac{\alpha_m \xi_m}{N}}, \end{aligned} \quad (70)$$

for $m = 1, \dots, N_t$, where \mathbf{c}_m is the m th column of $\sum_{i,j} \mathbf{B}_{i,j}$. Problem (70) is a nearest vector problem with low PAR constraint and has been well studied in [2] via Karush-Kuhn-Tucker (KKT) conditions. By using the well-developed algorithms in [2] to solve each problem (70), the overall algorithm is summarized in Algorithm 3. Note that Algorithm 3 shares the same convergence property as that of Algorithm 1. Furthermore, the acceleration scheme based on the SQUAREM method is also applicable here, and the procedure is similar to Algorithm 2.

Algorithm 3 Design of optimal training sequence for the MMSE minimization (26) or the CMI maximization (31) under the PAR constraint.

- 1: Set $t = 0$, and initialize $\mathbf{U}^{(0)}$ such that $\max_n \{|u_{n,m}^{(0)}|\} \leq \sqrt{\frac{\alpha_m}{N}}$, $m = 1, \dots, N_t$.
- 2: **repeat**
- 3: $\mathbf{S}^{(t)} = \mathcal{T}(\mathbf{U}^{(t)})$, and $\tilde{\mathbf{S}}^{(t)} = \mathbf{I}_{N_r} \otimes \mathbf{S}^{(t)}$
- 4: $\mathbf{A}^{(t)} = \left(\tilde{\mathbf{S}}^{(t)} \mathbf{R}_0 (\tilde{\mathbf{S}}^{(t)})^H + \mathbf{W} \right)^{-1} \tilde{\mathbf{S}}^{(t)} \mathbf{R}_0$
- 5: $\mathbf{V}^{(t)} = \begin{cases} \mathbf{I}, & \text{for the MMSE minimization} \\ \mathbf{R}^{(t)}, & \text{for the CMI maximization} \end{cases}$
- 6: $\lambda^{(t)} = \|\mathbf{R}_0\|_1 \left\| \mathbf{A}^{(t)} \mathbf{V}^{(t)} (\mathbf{A}^{(t)})^H \right\|_1$
- 7: $\mathbf{B}(\tilde{\mathbf{S}}^{(t)}, \mathbf{V}^{(t)}) = \lambda^{(t)} \tilde{\mathbf{S}}^{(t)} - \mathbf{A}^{(t)} \mathbf{V}^{(t)} (\mathbf{A}^{(t)})^H \tilde{\mathbf{S}}^{(t)} \mathbf{R}_0 + \mathbf{A}^{(t)} \mathbf{V}^{(t)} \mathbf{R}_0$
- 8: $\mathbf{U}_{:,m}^{(t+1)} \in \underset{\substack{\max_n \{|u_{n,m}|\} \leq \sqrt{\frac{\alpha_m \xi_m}{N}} \\ \|\mathbf{U}_{:,m}\|^2 = \alpha_m}}{\arg \min} \|\mathbf{U}_{:,m} - \mathbf{c}_m\|^2, m = 1, \dots, N_t$
- 9: $t \leftarrow t + 1$
- 10: **until** convergence

V. NUMERICAL EXAMPLES

In this section, we employ proposed algorithms to design unimodular and low PAR sequences for channel estimation. For SISO channels, we compare the channel estimation error and the obtained conditional mutual information of our proposed sequences with that of low sidelobe or random phases. For MIMO channel estimation, the same performance metrics are compared for our proposed sequences, sequences of good auto- and cross-correlation properties, and sequences of random phases. Then we show the advantage of optimal low PAR sequences over sequences of random phases in the MIMO channel estimation.

A. Unimodular Sequences for SISO Channel Estimation

In this subsection, numerical results are presented to illustrate the advantage of considering the prior information in the

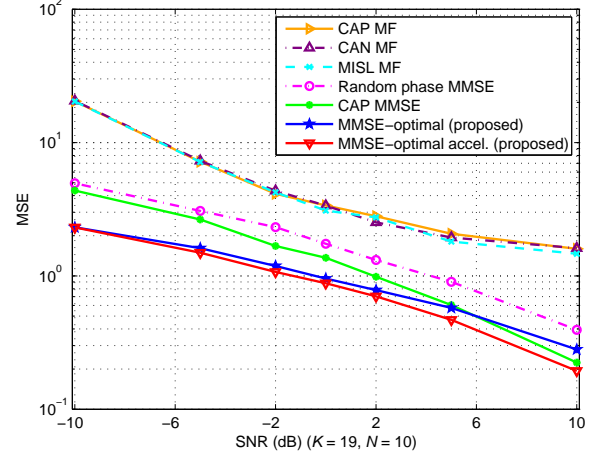


Fig. 1. MSE of SISO channel estimates with different unimodular training sequences. The results are averaged over 200 Monte Carlo simulations.

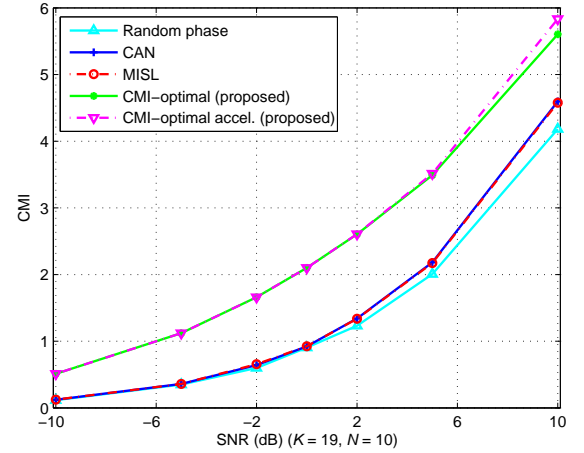


Fig. 2. The CMI with different unimodular training sequences for SISO channels. The results are averaged over 200 Monte Carlo simulations.

design of unimodular sequences for channel estimation and conditional mutual information maximization. Let $N_t = N_r = 1$, and we can apply Algorithm 1 and Algorithm 2 to design optimal unimodular training sequences for a SISO channel. We compute the MMSE estimates with our proposed sequences, sequences of low sidelobe, and sequences of random phase, and then compare the resulting MSE with matched filtering (MF) using low sidelobe sequences.

The underlying channel impulse response is chosen by $\mathbf{h}_{\text{true}} \sim \mathcal{CN}(\mathbf{0}_{K+1}, \mathbf{R}_{\text{true}})$ with length $K + 1 = 20$, and $(\mathbf{R}_{\text{true}})_{i,j} = 0.9^{|i-j|} 0.9^{\frac{i-1}{2}} 0.9^{\frac{j-1}{2}}$ for $i, j = 1, \dots, K+1$. The channel is thus correlated with exponentially decreasing power with respect to time delay, which corresponds to the correlated scattering environment with multipath fading in wireless communications [33]. The length of training sequence is $N = 10$. The channel noise is set to be $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}_{N+K}, \mathbf{W})$ with $(\mathbf{W})_{i,j} = 0.2^{|i-j|}$ for $i, j = 1, \dots, N + K$. Considering the inaccuracy of channel covariance matrix in hand, the optimal unimodular sequence \mathbf{u} is designed under the assumed prior

$\mathbf{h}_0 \sim \mathcal{CN}(\mathbf{0}_{K+1}, \mathbf{R}_0)$ and $(\mathbf{R}_0)_{i,j} = 0.8^{|i-j|} 0.8^{\frac{i-1}{2}} 0.8^{\frac{j-1}{2}}$. The mean square error (MSE) of the channel estimator is then

$$\text{MSE}(\hat{\mathbf{h}}_{\text{MMSE}}) = \|\hat{\mathbf{h}}_{\text{MMSE}} - \mathbf{h}_{\text{true}}\|_2^2, \quad (71)$$

where $\hat{\mathbf{h}}_{\text{MMSE}}$ is given by (21) and $\mathbf{S} = \mathcal{T}(\mathbf{u})$. Based on the true channel covariance matrix, the conditional mutual information obtained with training sequence \mathbf{u} is

$$\text{CMI}(\mathbf{u}) = \frac{1}{2} \log \det (\mathbf{I} + \mathbf{R}_{\text{true}} \mathbf{S}^H \mathbf{W}^{-1} \mathbf{S}). \quad (72)$$

The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{u}\|^2 / N}{\text{Tr}(\mathbf{W}) / (N + K)} \text{ (dB)}. \quad (73)$$

For different values of SNR, the resulting MSE and CMI are approximated by running 200 times Monte Carlo simulations. In our simulations, both Algorithm 1 and Algorithm 2 are initialized with unimodular sequences of random phases uniformly distributed in $[0, 2\pi]$. And the algorithms are considered to be converged when the difference between two consecutive updates is no larger than 10^{-6} , i.e., $\|\mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}\|_2 \leq 10^{-6}$.

Fig. 1 shows the MSE of different channel estimates after training with different unimodular sequences. Both CAP and CAN were proposed to design sequences with low sidelobes, or good correlation properties, and sequences designed by CAP was employed to estimate channel impulse response with the matched filter [5]. It was claimed that MISL could further reduce the sidelobes of the designed unimodular sequences [6], with which channel estimate by matched filtering was also compared herein. The resulting MSE of our proposed sequence, MMSE-optimal accel., by the accelerated scheme Algorithm 2 is lower than that of low sidelobes and that of random phases, especially in the low SNR scenarios. Therefore, the good correlation properties do not guarantee a good channel estimate when the length of the training sequence is limited with respect to the length of the channel impulse response. Note that sequence MMSE-optimal by Algorithm 1 achieves almost the same performance as that of MMSE-optimal accel., but the resulting MSE degrades a little bit in the high SNR case as it needs more iterations to converge. The convergence of Algorithm 1 and Algorithm 2 will be illustrated in Section V-D.

The obtained CMI for different unimodular sequences are shown in Fig. 2. Although by definition (72), the resulting CMI only depends on the channel statistics without being affected by the channel realizations, Monte Carlo simulations are still conducted for 200 times to avoid the effects from local minima. Expectedly, sequences obtained by CAN and MISL produces almost the same CMI. By incorporating the prior channel information into the sequence design, however, the CMI obtained is improved.

B. Unimodular Sequences for MIMO Channel Estimation

In this subsection, we compare the optimal unimodular sequences with those of good correlation properties [7] or random phases for MIMO channels. As in the case of SISO

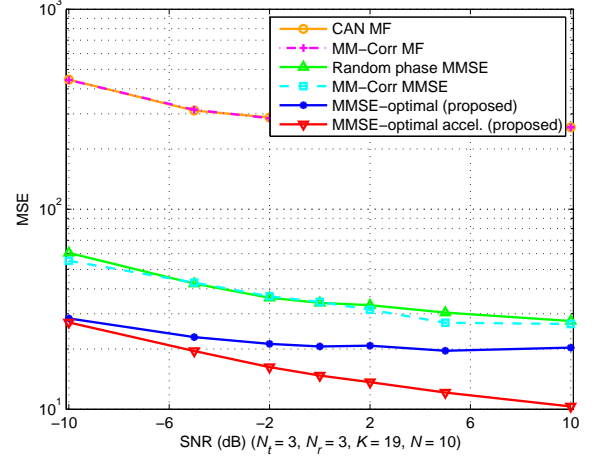


Fig. 3. MSE of MIMO channel estimates with different unimodular training sequences. The results are averaged over 100 Monte Carlo simulations.

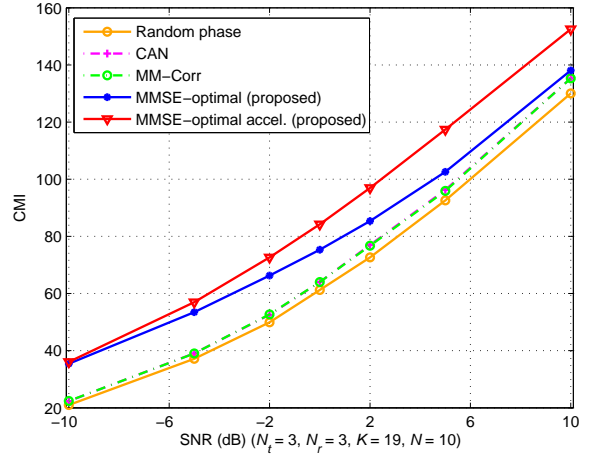


Fig. 4. The CMI with different unimodular training sequences for MIMO channels. The results are averaged over 100 Monte Carlo simulations.

channels, two performance metrics are considered, namely the channel MSE and CMI.

Suppose the MIMO channel has $N_t = 3$ transmit antennas and $N_r = 3$ receive antennas, with the length of the channel impulse $K + 1 = 20$. The vectorized channel impulse response \mathbf{h}_{true} is drawn from a circular complex Gaussian distribution $\mathcal{CN}(\mathbf{0}_{N_t N_r (K+1)}, \mathbf{R}_{\text{true}})$. Each channel coefficient $(\mathbf{h}_{\text{true}})_i, i = 1, \dots, N_t N_r (K + 1)$ is associated with a triple set (n_t, n_r, k) , where $n_t = 1, \dots, N_t$ and $n_r = 1, \dots, N_r$ are indices of transmit and receive antenna, respectively, and $k = 0, \dots, K$ is the channel delay. And each entry $(\mathbf{R}_{\text{true}})_{i,j}$ of the covariance matrix describes the correlation between the channel coefficient of the triple set (n_{t1}, n_{r1}, k_1) and (n_{t2}, n_{r2}, k_2) . Without loss of generality, consider

$$\mathbf{R}_{\text{true}} = \mathbf{R}_r \otimes \mathbf{R}_d \otimes \mathbf{R}_t \quad (74)$$

where $(\mathbf{R}_r)_{n_{r1}, n_{r2}} = \rho_1^{|n_{r1} - n_{r2}|}$ and $(\mathbf{R}_t)_{n_{t1}, n_{t2}} = \rho_3^{|n_{t1} - n_{t2}|}$ characterizes, respectively, the correlation between transmit antennas and the correlation between receive antennas, and

$(\mathbf{R}_d)_{k_1, k_2} = \rho_2^{|k_1 - k_2|}$ is an exponentially decaying correlation with respect to the channel delay. For the true channel impulse response \mathbf{h}_{true} , we set $\rho_1 = \rho_3 = 0.9$ and $\rho_2 = 0.7$. In the optimal unimodular training sequence design, the channel prior \mathbf{h}_0 is assumed to follow a circularly complex Gaussian distribution with zero mean and covariance matrix \mathbf{R}_0 of the same correlation structure as (74) and $\rho_1 = \rho_3 = 0.8$ and $\rho_2 = 0.6$. Each column of noise matrix \mathbf{V} in model (5) corresponds to a MISO channel, and the vectorized noise is assumed to be colored with a Toeplitz correlation and $\text{vec}(\mathbf{V}) \sim \mathcal{CN}(\mathbf{0}_{(N+K)N_t}, \mathbf{W})$, with $W_{i,j} = 0.2^{|i-j|}$, $i, j = 1, \dots, (N+K)N_r$. The optimal unimodular training sequences, sequences of good auto- and cross-correlations properties, and sequences of random phases are transmitted and then the corresponding MMSE channel estimators can be obtained. The MSE for each estimate is calculated by (71) with $\mathbf{S} = \mathcal{T}(\mathbf{U})$. The CMI is similarly defined by (72). The SNR is defined as

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{U}\|_F^2 / (NN_t)}{\text{Tr}(\mathbf{W}) / ((N+K)N_r)} \text{ (dB)}. \quad (75)$$

The setting for algorithm initialization and convergence are the same as the unimodular case. And the MSE and CMI are averaged over 100 times Monte Carlo simulations for different values of SNR.

Fig. 3 shows the MSE of MMSE channel estimates with different unimodular training sequences and SNR's. The length of sequence for each transmit antenna is $N = 10$. It is obvious that the optimal unimodular sequences, both MMSE-optimal by Algorithm 1 and MMSE-optimal accel. by Algorithm 2, produce smaller MSE than that of random phases or good auto- and cross-correlation properties (Good-Corr). Also notice that there is a gap between two curves of MSE of MMSE-optimal and MMSE-optimal accel. This is because Algorithm 1 needs much more iterations to be converged for MIMO channel training sequence design than that of the SISO case. The convergence properties are shown in Section V-D.

In the CMI maximization for MIMO channels, the performances of different unimodular sequences are shown in Fig. 4 with $N = 10$. For different SNR, the optimal unimodular training sequences can achieve larger CMI than sequences of either random phase or good correlation properties.

C. Low PAR Sequences for MIMO Channel Estimation

Consider the MIMO channel of the same conditions described in Section V-B. We employ Algorithm 3 and its accelerated scheme to design low PAR sequences for the application of MMSE channel estimation. In Fig. 5, MMSE-optimal and MMSE-optimal accel. are obtained by Algorithm 3 and its accelerated scheme, respectively. It is demonstrated that both optimal training sequences achieve much smaller MSE than low PAR sequences of random phases. Like the results for Algorithm 1 and Algorithm 2 in the previous subsections, MMSE-optimal renders an larger MSE than MMSE-optimal accel. especially in the high SNR cases. An example of convergence of both algorithms are shown in Section V-D. In Fig. 6, we also compare resulting MSE of unimodular sequences and sequences of different values of PAR.

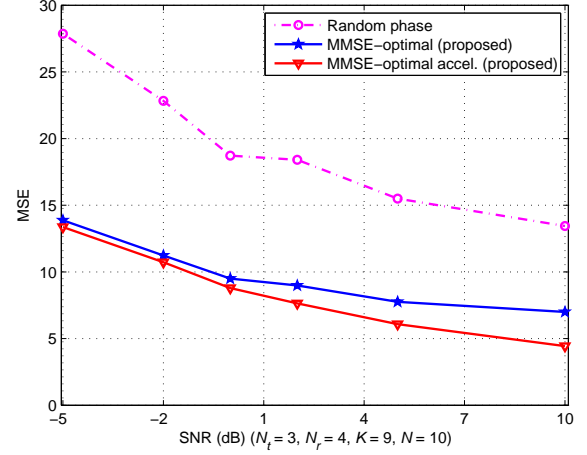


Fig. 5. MSE with different low PAR training sequences for MIMO channels. PAR = {1, 2, 3} with power proportions among three antennas: 1 : 2 : 3. The results are averaged over 100 Monte Carlo simulations.

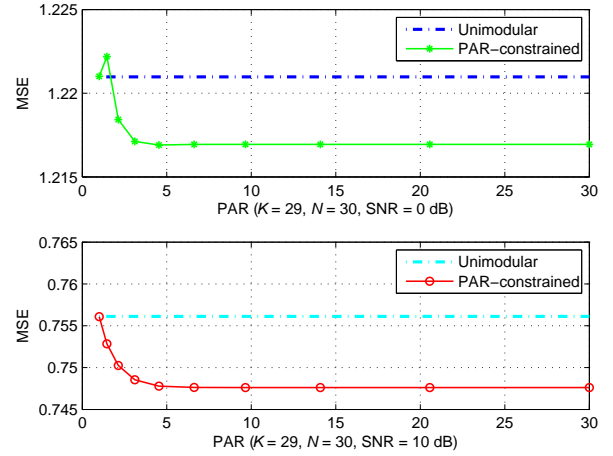


Fig. 6. MSE for SISO channel estimation with PAR-constrained sequences or unimodular sequences. The results are averaged over 100 Monte Carlo simulations.

D. Convergence of Proposed Algorithms

Experimental results are given to show the convergence properties of proposed algorithms for the MMSE minimization problem and the CMI maximization problem with unimodular constraints or low PAR constraints. The setting for algorithm initialization and convergence criteria are the same as previous subsections. First, we experiment with Algorithm 1 and Algorithm 2 for both MMSE minimization and CMI maximization in SISO channel unimodular training sequence design. Fig. 7 shows the objective values with respect to algorithm iterations. In both problems, Algorithm 1 converge monotonically to a stationary point though slowly. With acceleration techniques, however, Algorithm 2 renders an very fast convergence. The same convergence properties can be seen in Fig. 8, where unimodular sequences for MIMO channel estimation are considered with $N_t = 3$, $N_r = 4$. Within the same MIMO channel setting, Algorithm 3 and its accelerated scheme are

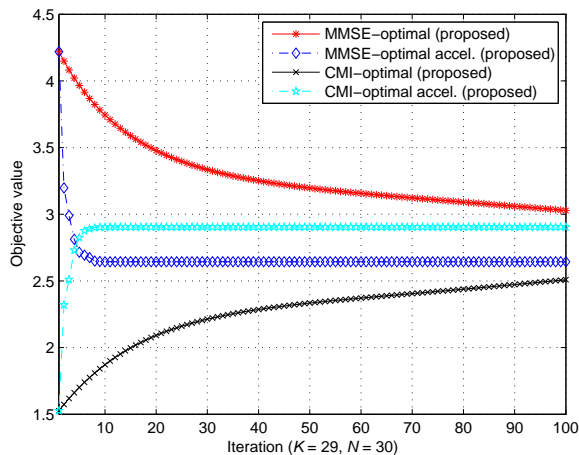


Fig. 7. Convergence of algorithms for optimal unimodular sequence design for SISO channel estimation, SNR = -5 dB.

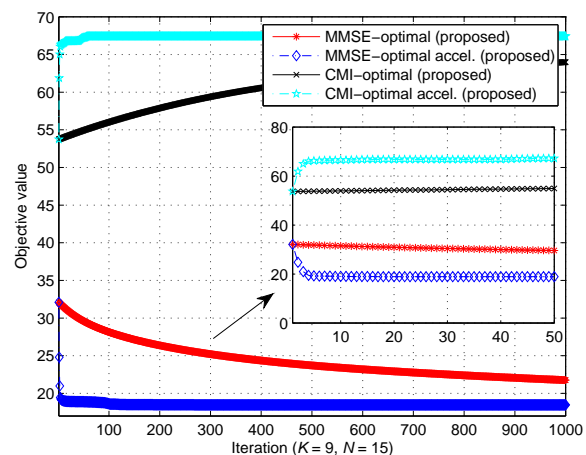


Fig. 8. Convergence of algorithms for optimal unimodular sequence design for MIMO channel estimation, $N_t = 3$, $N_r = 4$, and SNR = -5 dB.

applied to design low PAR sequences. The convergence of both algorithms are shown in Fig. 9. Note that in those three examples, the algorithms Algorithm 1 and Algorithm 3 converge slower than the accelerated scheme especially in designing sequences for MIMO channels with large values of SNR. This is due to successive majorizations or minorizations applied in the derivation of algorithms and thus explains the difference between two training sequences in terms of the resulting MSE and CMI.

VI. CONCLUSION

In this paper, optimal training sequences with unimodular constraint and low PAR constraints are considered. The optimal sequence design problem is formulated by minimizing the MMSE criterion and maximizing the CMI criterion. The formulated problems are nonconvex and efficient algorithms are developed based on the majorization-minimization framework. Furthermore, the acceleration scheme is derived using the SQUAREM method. All the proposed algorithms are

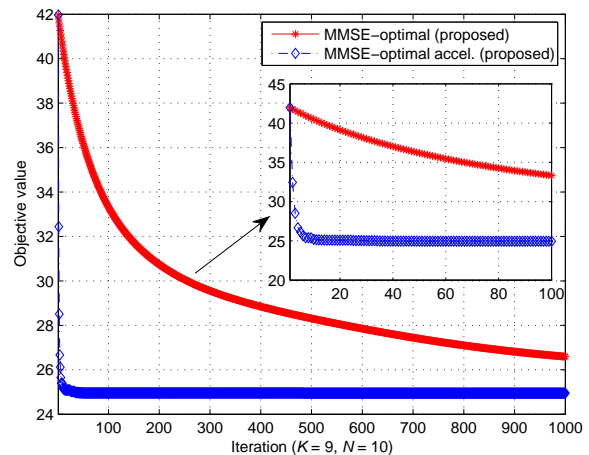


Fig. 9. Convergence of algorithms for optimal low PAR sequence design for MIMO channel estimation, $N_t = 3$, $N_r = 4$, and SNR = -5 dB.

guaranteed to monotonically converge to a stationary point. Numerical results show that the optimal unimodular sequences can improve either the accuracy of channel estimate or the CMI compared with those of sequences with good correlation properties or random phases. Under the same criteria, the optimal sequence design with low PAR constraint is also studied, for which the similar algorithms to unimodular case are derived. Numerical examples show that the optimal low PAR sequences perform better than that of random phases.

REFERENCES

- [1] S. Benedetto and E. Biglieri, *Principles of digital transmission: with wireless applications*. New York: Kluwer Academic, 1999.
- [2] J. Tropp, I. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan 2005.
- [3] H. He, P. Stoica, and J. Li, "Designing unimodular sequence sets with good correlations—including an application to MIMO radar," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4391–4405, Nov 2009.
- [4] —, "Wideband MIMO systems: Signal design for transmit beam-pattern synthesis," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 618–628, Feb 2011.
- [5] P. Stoica, H. He, and J. Li, "New algorithms for designing unimodular sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1415–1425, Jan. 2009.
- [6] J. Song, P. Babu, and D. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, Apr. 2015.
- [7] J. Song, P. Babu, and D. P. Palomar, "Sequence Set Design With Good Correlation Properties via Majorization-Minimization," *ArXiv*, Oct. 2015. [Online]. Available: <http://arxiv.org/abs/1510.01899>
- [8] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.
- [9] Y. Liu, T. F. Wong, and W. W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.
- [10] D. Katselis, E. Kofidis, and S. Theodoridis, "On training optimization for estimation of correlated MIMO channels in the presence of multiuser interference," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4892–4904, Oct. 2008.
- [11] M. Biguesh, S. Gazor, and M. H. Shariat, "Optimal training sequence for MIMO wireless systems in colored environments," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3144–3153, Mar. 2009.

- [12] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.
- [13] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
- [14] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Found. Trends Commun. Inf. Theory*, vol. 3, no. 4, pp. 331–551, Nov. 2006. [Online]. Available: <http://dx.doi.org/10.1561/01000000018>
- [15] D. Katselis, C. R. Rojas, M. Bengtsson, E. Björnson, X. Bombois, N. Shariati, M. Jansson, and H. Hjalmarsson, "Training sequence design for MIMO channels: an application-oriented approach," *arXiv preprint arXiv:1301.3708*, 2013.
- [16] M. Bengtsson, "Algorithmic solutions for pilot design optimization in arbitrarily correlated scenarios," in *2014 IEEE Int. Conf. Acoust. Speech and Signal Process. (ICASSP)*, 2014, pp. 5715–5719.
- [17] Q. Shi, C. Peng, W. Xu, and Y. Wang, "Training signal design for MIMO channel estimation with correlated disturbance," in *2014 IEEE Int. Conf. Acoust. Speech and Signal Process. (ICASSP)*, 2014, pp. 5720–5724.
- [18] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2572–2583, Sep. 2004.
- [19] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.
- [20] G. Leus and A.-J. van der Veen, "Optimal training for ML and LMMSE channel estimation in MIMO systems," in *2005 IEEE/SP 13th Workshop Statistical Signal Process.*, 2005, pp. 1354–1357.
- [21] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998, vol. 2.
- [22] J. H. Manton, "Optimal training sequences and pilot tones for OFDM systems," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 151–153, Apr. 2001.
- [23] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 2012.
- [24] M. Golay, "A class of finite binary sequences with alternate autocorrelation values equal to zero (corresp.)," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 449–450, May 1972.
- [25] N. Zhang and S. Golomb, "Polyphase sequence with low autocorrelations," *IEEE Trans. I*, vol. 39, no. 3, pp. 1085–1089, May 1993.
- [26] Y. Yang, R. Blum, Z. He, and D. Fuhrmann, "MIMO radar waveform design via alternating projection," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1440–1445, March 2010.
- [27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [28] T. N. Davidson, "Enriching the art of FIR filter design via convex optimization," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 89–101, May 2010.
- [29] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
- [30] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*, D. Palomar and Y. Eldar, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2009, pp. 42–88.
- [31] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [32] R. Varadhan and C. Roland, "Simple and globally convergent methods for accelerating the convergence of any EM algorithm," *Scand. J. Statist.*, vol. 35, no. 2, pp. 335–353, 2008.
- [33] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.