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

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

***Abstract*—Communication systems and radars widely employ sequences of low peak-to-average power ratio (PAR) or unimodulus to meet the hardware constraints and maximize the power efficiency. Numerous works have probed the unimodular sequence design especially attempting to obtain good correlation properties. Regarding channel estimation, however, sequences of such properties do not necessarily qualify for the mission. And tailored unimodular sequences for the specific criterion concerned are more desirable when we have access to the prior knowledge of the channel impulse response. In this paper, we formulate the problem of unimodular sequence design by optimizing minimum mean square error and conditional mutual information, respectively. The problems turn out nonconvex and we develop efficient algorithms based on the majorization–minimization framework with convergence guaranteed. More general, we also examine optimal sequence design with low PAR constraints. Numerical examples demonstrate the improved results of mean square error, signal-to-noise ratio, and conditional mutual information by using our proposed training sequences, with the efficiency of the derived algorithms illustrated.**

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

I. Introduction

Sequences with the peak-to-average power ratio (PAR) constraints find 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., unimodulus, to be transmitted [1]. In MIMO radars and code-division multiple-access (CDMA) applications, the practical implementation demands of hardware such as radio frequency power amplifiers and analog-to-digital converters require the transmitted sequences to be unimodular or of low PAR [2]–[4]. In this paper, we consider the design of unimodular and low PAR sequences for channel estimation.

There is an extensive literature on designing unimodular sequences with good correlation properties such that the autocorrelation of the sequence is zero at each nonzero lag. Whereas such properties are usually difficult to achieve, alternative metrics of "goodness" have been suggested where autocorrelation sidelobes are suppressed rather than exactly eliminated. Minimizing integrated sidelobe level (ISL) and maximizing ISL-related merit factor (MF) are put forward in [5] with cyclic algorithms (CA) proposed to solve the formulated optimization problems. In [6], a further computationally efficient algorithm called MISL for minimizing ISL is proposed, reporting lower autocorrelation sidelobes with less computational complexity. The more general weighted ISL is addressed in [5], [7].

The good correlation properties of a single unimodular sequence are also extended to a set of sequences so that the spatial diversity is further exploited. Apart from good autocorrelation properties for each sequence, the good cross-correlation requires each sequence to be nearly uncorrelated with time-shifted versions of the other sequences [3]. Some other numerical algorithms quest for refined sequence sets with enhanced good correlation properties [8]. Similarly, [9] explores methods to synthesize a desired beam pattern while maximally suppressing both the autocorrelation and cross-correlation sidelobes at/between given spacial angles.

The aforementioned ISL and ISL-related metrics indeed portray no more than 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 prior information of the channel impulse response as the unimodular low-ISL sequences are employed for the ensuing estimation. Such prior information, however, is usually available without incurring considerable expenses. Herein we will harness the potential of the second order channel statistics in the sequence design.

On the other hand, the unimodular constraint is a special case of the low PAR constraint, i.e., unit-PAR. The low PAR constraint, as a structural requirement, has been well studied in the design of tight frames [2]. Although we can adjust the individual vector norms of a frame to maximize, e.g., sum-capacity of communications links, [2] has not established the optimality regarding specific performance measures. Furthermore, the algorithm they proposed is based on alternating projection that often suffers a slow convergence rate. General optimization problems with quadratic objectives derived from, e.g., SNR maximization, have also been studied in [10], [11].

Within the PAR constraints, rather than exploit only the spatial diversity, we design sequences tailored for specific performance measures with channel covariance incorporated and expectedly

Manuscript received February 24, 2016; revised July 19, 2016; accepted August 28, 2016. Date of publication September 8, 2016; date of current version October 6, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hongbin Li. This work was supported in part by Hong Kong RGC under Grant 16206315 and in part by the Hong Kong RGC Theme-based Research Scheme under Grant T21-602/15R.

Z. Wang and D. P. Palomar are with the Hong Kong University of Science and Technology, Hong Kong (e-mail: zwangaq@ust.hk; palomar@ust.hk).

P. Babu was with the Hong Kong University of Science and Technology, Hong Kong. He is now with CARE, IIT Delhi, Delhi 110016, India (e-mail: prabhubabu@care.iitd.ac.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2607181

improve the quality of constructed estimators. As to sequence design for channel estimation, many studies have been conducted for both frequency-flat and frequency-selective fading channels with minimum mean square error (MMSE) minimization and conditional mutual information (CMI) maximization. Those optimization problems, however, address only the energy constraint without addressing the unimodular or low PAR constraints; see [12]–[19] and references therein on training sequence design for flat MIMO channels.

More related to our work is training sequence design for frequency-selective fading channels as high-speed transmission applications entail [20]. Optimal sequences design for the MMSE channel estimation has been studied; see, e.g., [21]. MIMO radar waveform design for extended target identification is investigated in [22] based on MMSE and CMI criteria assuming the covariance matrix of the target impulse response is known. By a different approach, MIMO radar waveforms are designed to maximize the SINR with statistics of target and/or clutter impulse response [23]. The low PAR constraints, however, are not considered in those optimization problems, which often results in a deteriorated performance compared with unoptimized low PAR reference waveforms [24]. In addition, a two-stage approach—obtaining an optimal waveform covariance matrix followed by shaping PAR-constrained waveforms accordingly—has been suggested [25]. Nevertheless, the optimality of the designed waveforms with respect to the MMSE or the CMI was not validated. Moreover, this approach is susceptible to the fact that the optimal waveform covariance matrix is often unattainable. Some other heuristic methods are also proposed to capitalize on, e.g., the similarity constraint given a good reference sequence [24] or spectral containment restrictions with pre-specified amplitudes [26], [27].

We formulate the problem of designing 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 on the prior channel and noise covariance matrices, the problems are also challenging even if only the power constraint is imposed. To tackle those issues, we employ the majorization-minimization (MM) technique to develop efficient algorithms. Note that [6]–[8] have used the MM method to devise algorithms for unimodular sequence design, yet the optimization problems addressed in this paper are formulated by different criteria, which has deployed the prior channel statistics to benefit the estimation. By rewriting the objective functions in a more appropriate way, we obtain the majorizing/minorizing functions for the 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 we provide an acceleration scheme 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 numerical examples demonstrate the superior performance of our proposed sequences in the resulting MSE and CMI; also, the output SNR has been improved, which justifies our account of prior channel information.

The rest of this paper is organized as follows. In Section II, we describe the channel model and then formulate the problems of optimal unimodular sequence design. 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. We present some numerical examples in Section V, and draw a conclusion 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 $\mathbf{I}_n$, where the subscript n denotes the size; in the cases where the subscript is undeclared, the size is implicit in the context. The superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ denote respectively transpose, conjugate transpose, and complex conjugate. By $\mathrm{vec}(\mathbf{X})$, the vector is formed by stacking the columns of $\mathbf{X}$. The Kronecker product is denoted by $\otimes$. And $\mathrm{E}(\cdot)$ takes the expectation of a random variable. $\mathrm{Tr}(\cdot)$ 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 $\mathbf{H}_0, \ldots, \mathbf{H}_K \in \mathbb{C}^{N_r \times N_t}$. 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 $n = 1, \ldots, N$. For simplicity, we still call this sequence of vectors as a sequence, which is denoted by $\mathbf{U} = [u_{n,m}] = [\,\mathbf{u}_1 \ \cdots \ \mathbf{u}_N\,]^T \in \mathbb{C}^{N \times N_t}$. Considering the unimodular constraint with energy budget $\mathrm{Tr}(\mathbf{U}^H\mathbf{U}) = \alpha$, we want to design $\mathbf{U} \in \mathcal{U}$, where

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

And the received sequence is given by

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

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

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

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

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

be a block Toeplitz convolution matrix with $[\mathbf{U}^T \quad \mathbf{0}]^T$ being the first block and remaining blocks are obtained by a downward circular shift of the previous block. Note that since $\mathrm{Tr}(\mathbf{U}^H\mathbf{U}) = \alpha$, then $\mathrm{Tr}(\mathbf{S}^H\mathbf{S}) = \alpha(K+1)$. $\mathbf{H} = [\mathbf{H}_0 \quad \cdots \quad \mathbf{H}_K]^T \in \mathbb{C}^{(K+1)N_t\times N_r}$ 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

$$\mathbf{Y} = \mathbf{SH} + \mathbf{V}. \tag{5}$$

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 $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, $\mathbf{h} = \mathrm{vec}(\mathbf{H})$, and $\mathbf{v} = \mathrm{vec}(\mathbf{V})$, and based on $\mathrm{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})\mathrm{vec}(\mathbf{Y})$, we have

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

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 estimation (MLE). Assume $\mathbf{h}$ in the channel model (6) is a constant (thus no prior on the channel imposed), and the vectorized noise follows a circularly complex Gaussian distribution, $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. The MLE of channel impulse is [28]

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

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

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

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

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

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

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

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

Lemma 1 ([30]): Let $\mathbf{X} \in \mathbb{C}^{M\times N}$ be such that $\mathrm{Tr}(\mathbf{X}^H\mathbf{X}) \leq \mu$ for some constant μ. The minimum of $\mathrm{Tr}((\mathbf{X}^H\mathbf{X})^{-1})$ 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 $(\mathrm{Tr}(\mathbf{S}^{\mathrm{H}}\mathbf{S}) = \alpha(K+1))$

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

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

$$\begin{aligned} \underset{\mathbf{U},\mathbf{S}}{\text{minimize}} \quad & \left\|\mathbf{S}^H\mathbf{S} - \tfrac{\alpha}{N_t}\mathbf{I}\right\|_F^2 \\ \text{subject to} \quad & \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U} \in \mathcal{U}. \end{aligned} \tag{14}$$

The optimal $\mathbf{S}$ satisfying (13) portrays an impulse-like correlation shape pursued in [3], [8], where the aperiodic cross-correlation is defined as

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

for $m_1, m_2 = 1, \ldots, N_t$ and lags $k = 0, \ldots, N-1$. 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 $k = -(N-1), \ldots, 0, \ldots, (N-1)$ are given by

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} r_{1,1}(k) & r_{1,2}(k) & \cdots & r_{1,N_t}(k) \\ r_{2,1}(k) & r_{2,2}(k) & \cdots & r_{2,N_t}(k) \\ \vdots & \vdots & \ddots & \vdots \\ r_{N_t,1}(k) & r_{N_t,2}(k) & \cdots & r_{N_t,N_t}(k) \end{bmatrix}, \tag{16}$$

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

$$\boldsymbol{\Sigma} = \mathbf{S}^H\mathbf{S}, \tag{17}$$

and then we have

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_{-1} & \cdots & \boldsymbol{\Sigma}_{-K} \\ \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_0 & \cdots & \boldsymbol{\Sigma}_{-(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_K & \boldsymbol{\Sigma}_{K-1} & \cdots & \boldsymbol{\Sigma}_0 \end{bmatrix}. \tag{18}$$

Note that $\boldsymbol{\Sigma}$ 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

$$\begin{aligned} \underset{\mathbf{U}}{\text{minimize}} \quad & (K+1)\left\|\boldsymbol{\Sigma}_0 - \tfrac{\alpha}{N_t}\mathbf{I}\right\|_F^2 \\ & + 2\sum_{k=1}^{K}(K+1-k)\left\|\boldsymbol{\Sigma}_k\right\|_F^2 \\ \text{subject to} \quad & \mathbf{U} \in \mathcal{U}. \end{aligned} \tag{19}$$

The objective function of (19) is indeed the weighted correlation minimization criterion within the lag interval $k = 0, \ldots, K$ [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

$$\begin{array}{ll}\underset{\mathbf{U}}{\text{minimize}} & \left\|\boldsymbol{\Sigma}_0 - \frac{\alpha}{N_t}\mathbf{I}\right\|_F^2 + 2\sum_{k=1}^{N-1}\|\boldsymbol{\Sigma}_k\|_F^2 \\ \text{subject to} & \mathbf{U} \in \mathcal{U},\end{array} \tag{20}$$

for which an algorithm called CAN was developed in [5], and WeCAN can be employed as well. In [8], 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 $\mathrm{E}\{\|\hat{\mathbf{h}}_{\mathrm{MMSE}} - \mathbf{h}\|^2\}$, the MMSE estimator of the channel impulse $\mathbf{h}$ is given by

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

where $\tilde{\mathbf{S}} = \mathbf{I}_{N_r} \otimes \mathbf{S}$ [28]. And the error covariance matrix is

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

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

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

where the last equality is due to the matrix inversion lemma [29]. The MMSE is thus given by

$$\mathrm{MMSE}(\mathbf{S}) = \mathrm{Tr}(\mathbf{R}), \tag{25}$$

and the following problem can be formulated

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

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., [12]. The CMI is defined as

$$\mathrm{CMI}(\mathbf{S}) = I(\mathbf{h}; \mathbf{y}\,|\mathbf{S}) \tag{27}$$

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

where $H(\cdot)$ is the differential entropy of a distribution [31]. 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 $\mathrm{CMI}(\mathbf{S})$ can be written as

$$\mathrm{CMI}(\mathbf{S}) = \frac{1}{2}\log\left((2\pi e)^{(K+1)N_tN_r}\det(\mathbf{R}_0)\right) - \frac{1}{2}\log\left((2\pi e)^{(K+1)N_tN_r}\det(\mathbf{R})\right) \tag{29}$$

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

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

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

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., [32], [33], 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 [22] 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

$$\begin{aligned}\mathrm{MMSE}(\mathbf{S}) &= \mathrm{Tr}\left(\left(\mathbf{R}_0^{-1} + \frac{1}{\sigma^2}\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}\right)^{-1}\right) &(32)\\ &= \mathrm{Tr}\left(\left(\mathbf{R}_0^{-1} + \frac{1}{\sigma^2}\mathbf{I}_{N_r}\otimes\boldsymbol{\Sigma}\right)^{-1}\right), &(33)\end{aligned}$$

where the second equality follows from substitution of the correlation matrix $\boldsymbol{\Sigma} = \mathbf{S}^H\mathbf{S}$. Consider only the constraint $\mathrm{Tr}(\boldsymbol{\Sigma}) = (K+1)\alpha$ induced by the energy budget in (26), minimizing (33) with respect to $\boldsymbol{\Sigma}$ (instead of $\mathbf{S}$) can be rewritten as an SDP by resorting to the Schur-complement theorem [34], which yields the optimal correlation matrix $\boldsymbol{\Sigma}^\star$. Once $\boldsymbol{\Sigma}^\star$ is obtained, the problem boils down to recovering sequences from its correlation matrix, which is to solve the following approximation problem

$$\begin{array}{ll}\underset{\mathbf{U},\mathbf{S}}{\text{minimize}} & \|\mathbf{S}^H\mathbf{S} - \boldsymbol{\Sigma}^\star\|_F\\ \text{subject to} & \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U}\in\mathcal{U},\end{array} \qquad (34)$$

if zero error is achievable. For a single sequence and without the unimodular constraint, problem (34) can be tackled by means of filter design [35]. However, constructing a unimodular sequence that presents a prescribed correlation shape is challenging. As a special case, [3] and [8] 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 [36], [37].

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

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

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

$$g\left(\mathbf{x}^{(t)},\mathbf{x}^{(t)}\right) = f\left(\mathbf{x}^{(t)}\right), \qquad (36)$$

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

and

$$\mathbf{x}^{(t+1)} \in \arg\min_{\mathbf{x}\in\mathcal{X}} g\left(\mathbf{x},\mathbf{x}^{(t)}\right). \qquad (38)$$

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

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

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 [36].

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

$$\mathrm{MMSE}(\mathbf{S}) = \mathrm{Tr}\left(\mathbf{R}_0 - \mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{P}^{-1}\tilde{\mathbf{S}}\mathbf{R}_0\right). \qquad (40)$$

Lemma 2: The function $f(\mathbf{X},\mathbf{Z}) = \mathrm{Tr}(\mathbf{X}^H\mathbf{Z}^{-1}\mathbf{X})$ is a matrix fractional function and is jointly convex in $\mathbf{Z}\succ 0$ and $\mathbf{X}$ [34, pp. 108–111].

By Lemma 2, $\mathrm{MMSE}(\mathbf{S}) = \mathrm{Tr}(\mathbf{R}_0 - \mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{P}^{-1}\tilde{\mathbf{S}}\mathbf{R}_0)$ 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, $\mathrm{MMSE}(\mathbf{S})$ can be majorized as follows:

$$\begin{aligned}\mathrm{MMSE}(\mathbf{S}) &\leq g_{\mathrm{MMSE}}\left(\mathbf{S},\mathbf{S}^{(t)}\right) &(41)\\ &= \mathrm{MMSE}\left(\mathbf{S}^{(t)}\right) + \mathrm{Tr}\left(\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{A}^{(t)}\right)\\ &\quad - 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{R}_0\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\right)\right\}, &(42)\end{aligned}$$

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

$$\begin{array}{ll}\underset{\mathbf{U},\mathbf{S}}{\text{minimize}} & g_{\mathrm{MMSE}}\left(\mathbf{S},\mathbf{S}^{(t)}\right)\\ \text{subject to} & \mathbf{S} = \mathcal{T}(\mathbf{U}), \mathbf{U}\in\mathcal{U},\end{array} \qquad (43)$$

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

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

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 $\mathbf{Z}\succ 0$ [34, pp. 108–111]. 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, $\mathrm{CMI}(\mathbf{S})$ is jointly convex in $\{\tilde{\mathbf{S}},\mathbf{P}\}$, and we can obtain the following minorization

$$\mathrm{CMI}(\mathbf{S}) \geq g_{\mathrm{CMI}}\left(\mathbf{S},\mathbf{S}^{(t)}\right) \tag{45}$$
$$= \mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{R}_0\left(\mathbf{R}^{(t)}\right)^{-1}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\right)\right\} \tag{46}$$
$$= -\frac{1}{2}\mathrm{Tr}\left(\left(\mathbf{R}^{(t)}\right)^{-1}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{A}^{(t)}\right) + \mathrm{CMI}\left(\mathbf{S}^{(t)}\right), \tag{47}$$

where

$$\mathbf{R}^{(t)} = \mathbf{R}_0 - \mathbf{R}_0\left(\tilde{\mathbf{S}}^{(t)}\right)^H\left(\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\left(\tilde{\mathbf{S}}^{(t)}\right)^H+\mathbf{W}\right)^{-1}\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0 \tag{48}$$
$$= \mathbf{R}_0^{-1} + \left(\tilde{\mathbf{S}}^{(t)}\right)^H\mathbf{W}^{-1}\tilde{\mathbf{S}}^{(t)}. \tag{49}$$

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

$$\begin{array}{ll}\underset{\mathbf{U},\mathbf{S}}{\text{maximize}} & g_{\mathrm{CMI}}\left(\mathbf{S},\mathbf{S}^{(t)}\right)\\ \text{subject to} & \mathbf{S}=\mathcal{T}(\mathbf{U}),\,\mathbf{U}\in\mathcal{U}.\end{array} \tag{50}$$

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

$$g\left(\mathbf{S};\mathbf{S}^{(t)},\mathbf{V}^{(t)}\right) = \mathrm{Tr}\left(\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\mathbf{R}_0\tilde{\mathbf{S}}^H\mathbf{A}^{(t)}\right) - 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{R}_0\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\right)\right\}, \tag{51}$$

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

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

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 $\mathrm{Tr}(\mathbf{Z}\mathbf{X}\mathbf{M}\mathbf{X}^H)$ can be majorized by $-2\mathrm{Re}\{\mathrm{Tr}((\lambda\mathbf{X}^{(t)}-\mathbf{Z}\mathbf{X}^{(t)}\mathbf{M})^H\mathbf{X})\}+\lambda\|\mathbf{X}\|_F^2+\mathrm{const}$, 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

$$\mathrm{Tr}\left(\mathbf{Z}\mathbf{X}\mathbf{M}\mathbf{X}^H\right)$$
$$= \mathrm{vec}^H(\mathbf{X})\,\mathrm{vec}(\mathbf{Z}\mathbf{X}\mathbf{M}) \tag{53}$$
$$= \mathrm{vec}^H(\mathbf{X})\left(\mathbf{M}^T\otimes\mathbf{Z}\right)\mathrm{vec}(\mathbf{X}) \tag{54}$$
$$\leq -2\mathrm{Re}\left\{\mathrm{vec}^H(\mathbf{X})\left(\lambda\mathbf{I}-\mathbf{M}^T\otimes\mathbf{Z}\right)\mathrm{vec}\left(\mathbf{X}^{(t)}\right)\right\} + \mathrm{vec}^H\left(\mathbf{X}^{(t)}\right)\left(\lambda\mathbf{I}-\mathbf{M}^T\otimes\mathbf{Z}\right)\mathrm{vec}\left(\mathbf{X}^{(t)}\right) + \lambda\mathrm{vec}^H(\mathbf{X})\,\mathrm{vec}(\mathbf{X}) \tag{55}$$
$$= -2\mathrm{Re}\left\{\mathrm{Tr}\left(\lambda\mathbf{X}^H\mathbf{X}^{(t)}-\mathbf{Z}\mathbf{X}^{(t)}\mathbf{M}\mathbf{X}^H\right)\right\} + \lambda\|\mathbf{X}\|_F^2 + \mathrm{vec}^H\left(\mathbf{X}^{(t)}\right)\left(\lambda\mathbf{I}-\mathbf{M}^T\otimes\mathbf{Z}\right)\mathrm{vec}\left(\mathbf{X}^{(t)}\right). \tag{56}$$

Notice that the third term of the last equation is simply a constant. And a scalar version of Lemma 4 can be found in [8, Lemma 1]. ∎

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

$$g\left(\mathbf{S};\mathbf{S}^{(t)},\mathbf{V}^{(t)}\right)$$
$$\leq -2\mathrm{Re}\left\{\mathrm{Tr}\left(\lambda^{(t)}\tilde{\mathbf{S}}^H\tilde{\mathbf{S}}^{(t)}-\mathbf{A}^{(t)}\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\tilde{\mathbf{S}}^H\right)\right\} - 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{R}_0\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}\right)\right\} + \mathrm{const} \tag{57}$$
$$= -2\mathrm{Re}\left\{\mathrm{Tr}\left(\left(\lambda^{(t)}\tilde{\mathbf{S}}^{(t)}-\mathbf{A}^{(t)}\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0 + \mathbf{A}^{(t)}\mathbf{V}^{(t)}\mathbf{R}_0\right)^H\tilde{\mathbf{S}}\right)\right\} + \mathrm{const}, \tag{58}$$

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

$$\lambda_{\max}(\mathbf{R}_0) \leq \|\mathbf{R}_0\|_1, \tag{59}$$
$$\lambda_{\max}\left(\mathbf{A}^{(t)}\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\right) \leq \left\|\mathbf{A}^{(t)}\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\right\|_1, \tag{60}$$

where $\|\cdot\|_1$ is maximum column sum matrix norm [38]. With $\lambda_{\max}(\mathbf{X}\otimes\mathbf{Z})=\lambda_{\max}(\mathbf{X})\lambda_{\max}(\mathbf{Z})$, we propose

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

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

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

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 $N_t(K+1)(i-1)+N_t j$, for $i=1,\ldots,N_r$, and $j=1,\ldots,K+1$. To find the next update $\mathbf{U}^{(t+1)}$, note that (62) can be equivalently written as

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

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

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

where $\arg(\cdot)$ is taken element-wise.

The whole procedure is summarized in Algorithm 1. The iterations are deemed to be converged, e.g., when the difference between two consecutive updates for $\mathbf{U}$ is no larger than some admitted threshold. The Algorithm 1 mainly involves a matrix inverse operation and several matrix multiplications in each iteration. We assume $(N+K)N_r > (K+1)N_r N_t$, thus the number of entries in the received sequences is greater than that of channel coefficients. With the Gaussian eliminations, the per iteration computational complexity is in the order of $\mathcal{O}((N+K)^3 N_r^3)$.

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 $\mathbf{U}^{(t)}, t=0,1,\ldots$, 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 $f:\mathbb{R}^N \to \mathbb{R}$ be a smooth function. A point $\mathbf{x}^\star$ is a local minimum of f within a subset $\mathcal{X} \subset \mathbb{R}^N$ if

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

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

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 $\{\mathbf{U}^{(t)}, t=0,\ldots\}$ 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]. ∎

Algorithm 1: Design of Unimodular Training Sequence for the MMSE Minimization (26) or the CMI Maximization (31).

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

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], [8], we employ an off-the-shelf method, called squared iterative method (SQUAREM) [39], 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 [39]. 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 $\mathrm{MMupdate}(\mathbf{U}^{(t)})$. 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. The main computational cost comes from the two MM updates in each iteration, whose complexity is in the order of $\mathcal{O}((N+K)^3 N_r^3)$. Empirically, the accelerated algorithm converges to an acceptable solution within ten iterations, which is much faster than Algorithm 1.

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,

Algorithm 2: Accelerated Scheme for Designing Optimal Unimodular Training Sequence for the MMSE Estimation.

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

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; see [2], [29] and references therein. 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 mth antenna, for $m = 1, \ldots, N_t$. And PAR is usually defined for each sequence transmitted by a single antenna as

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

provided that the training energy for the mth 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 $1 \le \text{PAR}(\mathbf{U}_{:,m}) \le N$. When $\text{PAR}(\mathbf{U}_{:,m}) = 1$, PAR constraint reduces to the unimodular constraint. Given the PAR constraints for each transmit antenna

$$\text{PAR}(\mathbf{U}_{:,m}) \le \xi_m, m = 1, \ldots, N_t \tag{67}$$

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

$$\begin{array}{ll} \underset{\mathbf{U},\mathbf{S}}{\text{minimize}} & \text{MMSE}\left(\mathbf{S}\right) \\ \text{subject to} & \mathbf{S} = \mathcal{T}\left(\mathbf{U}\right), \\ & \|\mathbf{U}_{:,m}\|^2 = \alpha_m, \\ & \max_n\{|u_{n,m}|\} \le \sqrt{\frac{\alpha_m \xi_m}{N}}, m = 1, \ldots, N_t \end{array} \tag{68}$$

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)

$$\begin{array}{ll} \underset{\mathbf{U}}{\text{minimize}} & \left\|\mathbf{U} - \sum_{i,j}\mathbf{B}_{i,j}\right\|_F^2 \\ \text{subject to} & \|\mathbf{U}_{:,m}\|^2 = \alpha_m, \\ & \max_n\{|u_{n,m}|\} \le \sqrt{\frac{\alpha_m \xi_m}{N}}, m = 1, \ldots, N_t \end{array} \tag{69}$$

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

$$\begin{array}{ll} \underset{\mathbf{U}_{:,m}}{\text{minimize}} & \|\mathbf{U}_{:,m} - \mathbf{c}_m\|^2 \\ \text{subject to} & \|\mathbf{U}_{:,m}\|^2 = \alpha_m, \\ & \max_n\{|u_{n,m}|\} \le \sqrt{\frac{\alpha_m \xi_m}{N}}, \end{array} \tag{70}$$

for $m = 1, \ldots, N_t$, where $\mathbf{c}_m$ is the mth 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. The additional computational cost of Algorithm 3 is incurred by solving N_t nearest vector problems in each iteration. Specifically, solving each problem requires a sorting operation of computational complexity $\mathcal{O}(N\log N)$ and inner loops with worst case computational complexity $\mathcal{O}(N^2)$; and the number of inner loops varies from a few dozens to one hundred. For details, please refer to [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 $\max_n \{|u_{n,m}^{(0)}|\} \le \sqrt{\frac{\alpha_m}{N}}, m = 1, \ldots, N_t$.
2: **repeat**
3: $\quad \mathbf{S}^{(t)} = \mathcal{T}\left(\mathbf{U}^{(t)}\right)$, and $\tilde{\mathbf{S}}^{(t)} = \mathbf{I}_{N_r} \otimes \mathbf{S}^{(t)}$
4: $\quad \mathbf{A}^{(t)} = \left(\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0\left(\tilde{\mathbf{S}}^{(t)}\right)^H + \mathbf{W}\right)^{-1}\tilde{\mathbf{S}}^{(t)}\mathbf{R}_0$
5: $\quad \mathbf{V}^{(t)} = \begin{cases} \mathbf{I}, & \text{for the MMSE minimization} \\ \mathbf{R}^{(t)}, & \text{for the CMI maximization} \end{cases}$
6: $\quad \lambda^{(t)} = \|\mathbf{R}_0\|_1 \left\|\mathbf{A}^{(t)}\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H\right\|_1$
7: $\quad \mathbf{B}\left(\tilde{\mathbf{S}}^{(t)}, \mathbf{V}^{(t)}\right) = \lambda^{(t)}\tilde{\mathbf{S}}^{(t)} - \mathbf{A}^{(t)}\mathbf{V}^{(t)}\left(\mathbf{A}^{(t)}\right)^H \tilde{\mathbf{S}}^{(t)}\mathbf{R}_0 + \mathbf{A}^{(t)}\mathbf{V}^{(t)}\mathbf{R}_0$
8: $\quad \mathbf{U}_{:,m}^{(t+1)} \in \underset{\substack{\max_n\{|u_{n,m}|\} \le \sqrt{\frac{\alpha_m \xi_m}{N}} \\ \|\mathbf{U}_{:,m}\|^2 = \alpha_m}}{\arg\min} \|\mathbf{U}_{:,m} - \mathbf{c}_m\|^2$, $m = 1, \ldots, N_t$
9: $\quad t \leftarrow t+1$
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 MSE, output SNR, and the CMI of our proposed sequences with that of low sidelobes or random phases. For MIMO channel estimation, the MSE and CMI are compared among 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 unimodular ones 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 design of unimodular sequences for channel estimation and CMI 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 sidelobes, and sequences of random phase, and then compare the resulting MSE with matched filtering (MF) using low sidelobes sequences.

The true 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 $i, j = 1, \ldots, K+1$. 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 [40]. 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 $i, j = 1, \ldots, N+K$. 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

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

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

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

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

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

The output SNR by MMSE estimation is given by

$$\text{SNR}_{\text{out}} = 10 \log_{10} \frac{\|\hat{\mathbf{h}}_{\text{MMSE}} - \mathbf{F}\mathbf{v}\|^2}{\|\mathbf{F}\mathbf{v}\|^2} \text{(dB)}, \tag{74}$$

where $\mathbf{F} = \mathbf{R}_0 \tilde{\mathbf{S}}^H (\tilde{\mathbf{S}} \mathbf{R}_0 \tilde{\mathbf{S}}^H + \mathbf{W})^{-1}$. 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 $[0, 2\pi]$.

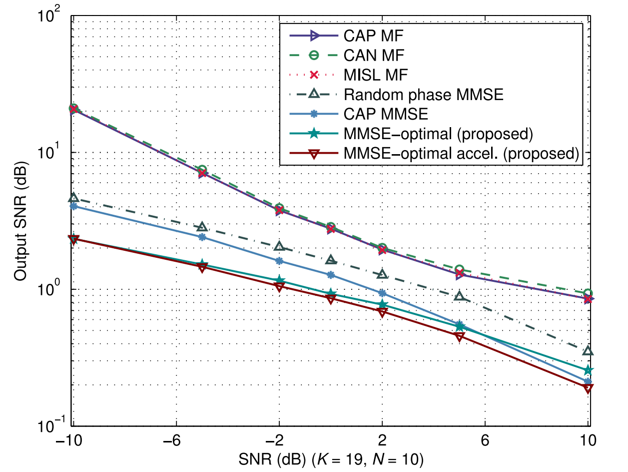

Fig. 1. MSE of SISO channel estimates with different unimodular training sequences. The results are averaged over 200 Monte Carlo simulations.

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 by 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. The resulting MSE of 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 output SNR with MMSE estimation and matched filters are also compared in Fig. 2, showing the improvement of our proposed sequences.

The obtained CMI for different unimodular sequences are shown in Fig. 3. 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 [8] or random phases for MIMO channels. As in the case of SISO channels,

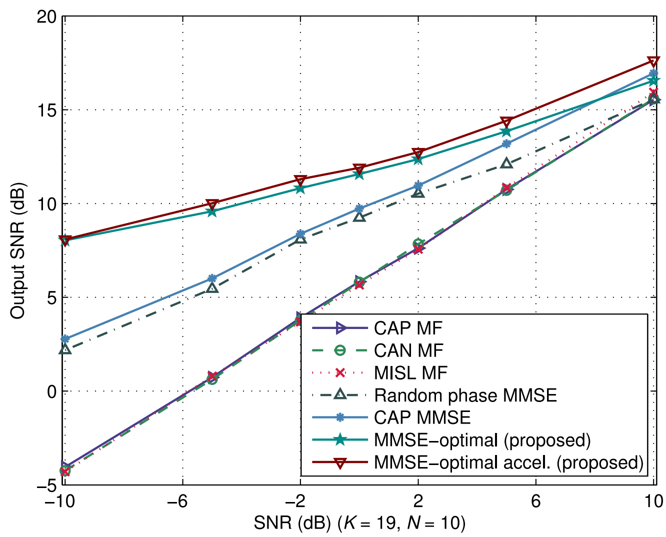

Fig. 2. Output SNR of SISO channel estimates with different unimodular training sequences. The results are averaged over 200 Monte Carlo simulations.

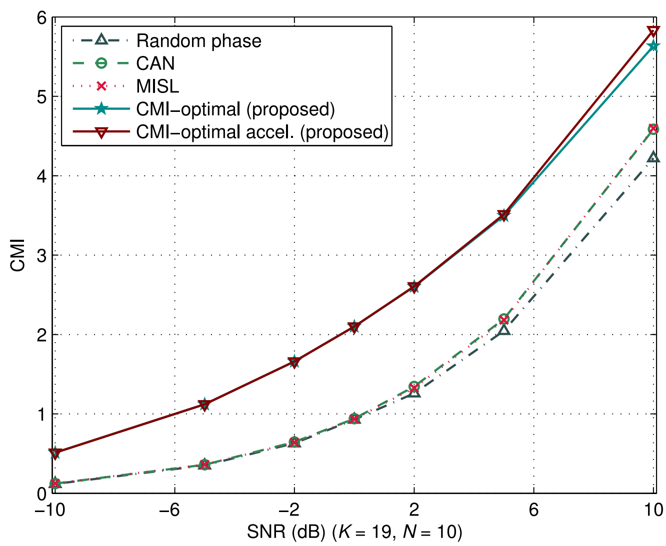

Fig. 3. The CMI with different unimodular training sequences for SISO channels. The results are averaged over 200 Monte Carlo simulations.

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 $\mathbf{h}_{\mathrm{true}}$ is drawn from a circular complex Gaussian distribution $\mathcal{CN}(\mathbf{0}_{N_t N_r (K+1)}, \mathbf{R}_{\mathrm{true}})$. Each channel coefficient $(\mathbf{h}_{\mathrm{true}})_i, i = 1, \ldots, N_t N_r(K+1)$ is associated with a triple set (n_t, n_r, k), where $n_t = 1, \ldots, N_t$ and $n_r = 1, \ldots, N_r$ are indices of transmit and receive antenna, respectively, and $k = 0, \ldots, K$ is the channel delay. And each entry $(\mathbf{R}_{\mathrm{true}})_{i,j}$ 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

$$\mathbf{R}_{\mathrm{true}} = \mathbf{R}_r \otimes \mathbf{R}_d \otimes \mathbf{R}_t \tag{75}$$

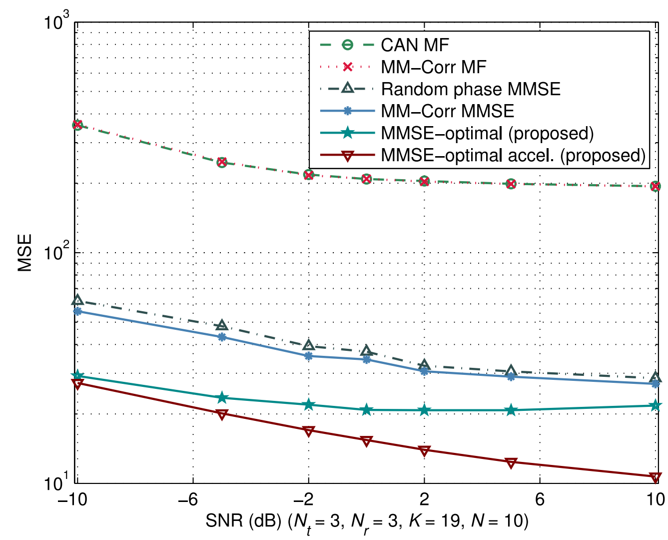

Fig. 4. MSE of MIMO channel estimates with different unimodular training sequences. The results are averaged over 100 Monte Carlo simulations.

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 $\mathbf{h}_{\mathrm{true}}$, 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 (75) 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 $\mathrm{vec}(\mathbf{V}) \sim \mathcal{CN}(\mathbf{0}_{(N+K)N_t}, \mathbf{W})$, with $W_{i,j} = 0.2^{|i-j|}, i, j = 1, \ldots, (N + K)N_r$. 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

$$\mathrm{SNR} = 10 \log_{10} \frac{\|\mathbf{U}\|_F^2 / (N N_t)}{\mathrm{Tr}(\mathbf{W}) / ((N + K) N_r)} \, (\mathrm{dB}). \tag{76}$$

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. 4 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

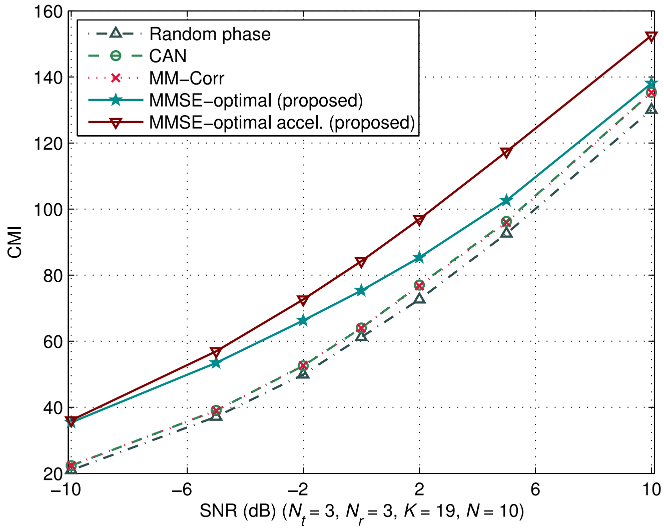

Fig. 5. The CMI with different unimodular training sequences for MIMO channels. The results are averaged over 100 Monte Carlo simulations.

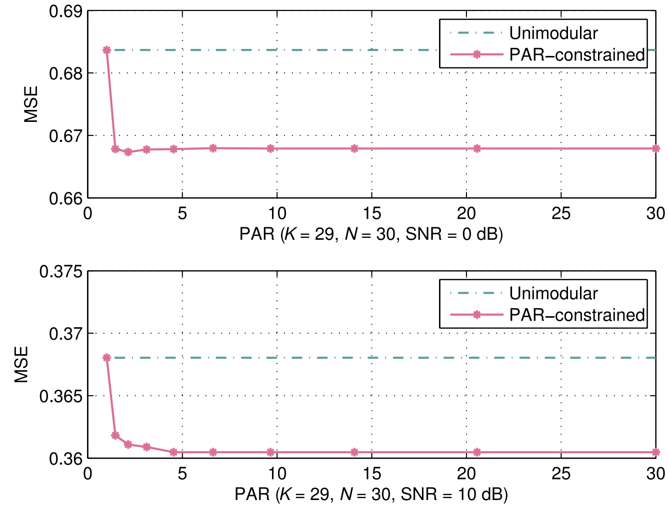

Fig. 7. MSE for SISO channel estimation with PAR-constrained sequences or unimodular sequences. The results are averaged over 100 Monte Carlo simulations.

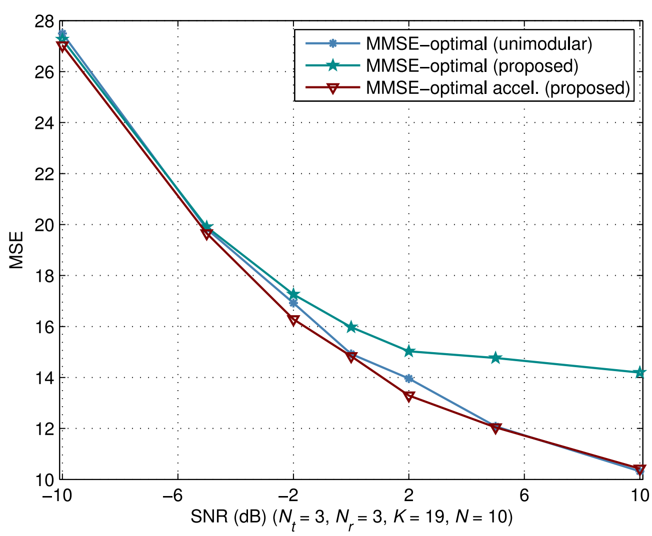

Fig. 6. MSE with different low PAR training sequences for MIMO channels. $\mathrm{PAR} = \{5, 5, 5\}$ with equal power for three antennas. The results are averaged over 100 Monte Carlo simulations.

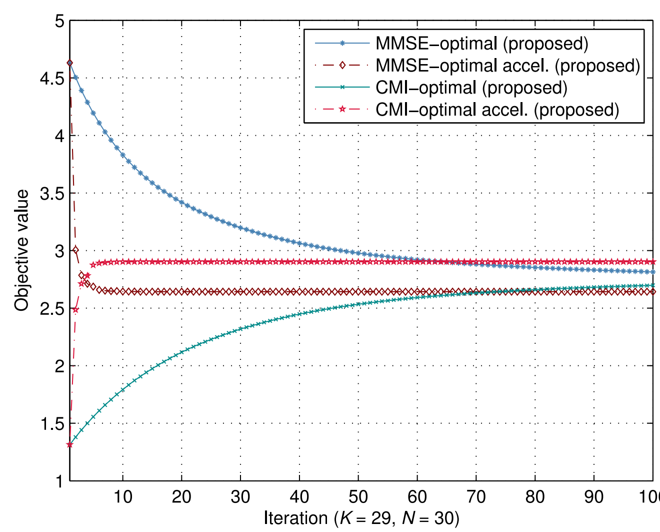

Fig. 8. Convergence of algorithms for optimal unimodular sequence design for SISO channel estimation, $\mathrm{SNR} = -5$ dB.

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. 5 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 MMSE channel estimation. In Fig. 6, 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 unimodular sequences. Like the results for Algorithm 1 and 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. Fig. 7 shows resulting MSE of sequences with different values of PAR.

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

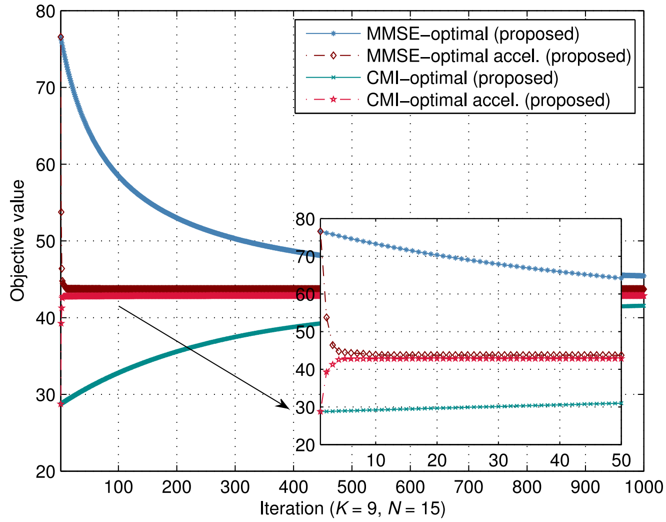

Fig. 9. Convergence of algorithms for optimal unimodular sequence design for MIMO channel estimation, $N_t = 3$, $N_r = 3$, and $\mathrm{SNR} = -5$ dB.

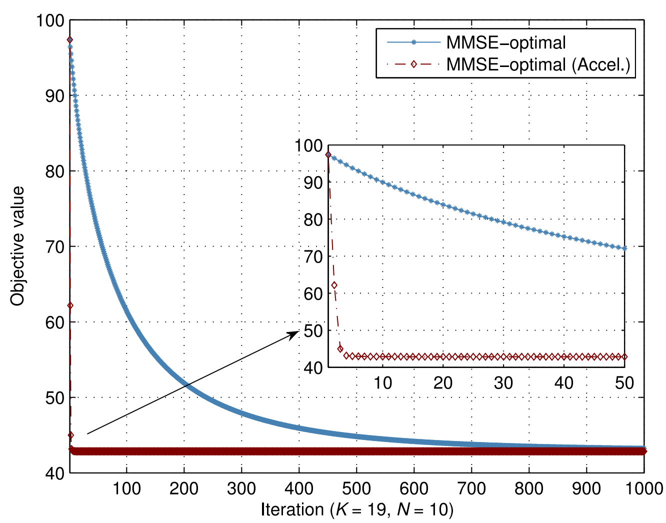

Fig. 10. Convergence of algorithms for optimal low PAR sequence design for MIMO channel estimation, $N_t = 3$, $N_r = 3$, and $\mathrm{SNR} = -5$ dB.

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. 8 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. 9, where unimodular sequences for MIMO channel estimation are considered with $N_t = 3$, $N_r = 3$. Within the same MIMO channel setting, Algorithm 3 and its accelerated scheme are applied to design low PAR sequences. The convergence of both algorithms are shown in Fig. 10. 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 and low PAR constraints are considered. The optimal sequence design problem is formulated by minimizing the MMSE criterion or 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 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 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 can improve on the unimodular sequences.

References

[1] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. New York, NY, USA: 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] H. He, P. Stoica, and J. Li, "Wideband MIMO systems: Signal design for transmit beampattern 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. Palomar, "Sequence design to minimize the weighted integrated and peak sidelobe levels," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2051–2064, Apr. 2016.
[8] J. Song, P. Babu, and D. Palomar, "Sequence set design with good correlation properties via majorization-minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2866–2879, Jun. 2016.
[9] Y. C. Wang, X. Wang, H. Liu, and Z. Q. Luo, "On the design of constant modulus probing signals for MIMO radar," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4432–4438, Sep. 2012.
[10] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1221–1234, Mar. 2014.
[11] A. De Maio, Y. Huang, M. Piezzo, S. Zhang, and A. Farina, "Design of optimized radar codes with a peak to average power ratio constraint," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2683–2697, Jun. 2011.
[12] 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.
[13] 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.
[14] 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.

[15] 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.
[16] 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.
[17] D. Katselis *et al.*, "Training sequence design for MIMO channels: An application-oriented approach," *EURASIP J. Wireless Commun. Netw.*, vol. 2013, no. 1, pp. 1–22, Oct. 2013.
[18] M. Bengtsson, "Algorithmic solutions for pilot design optimization in arbitrarily correlated scenarios," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 5715–5719.
[19] Q. Shi, C. Peng, W. Xu, and Y. Wang, "Training signal design for MIMO channel estimation with correlated disturbance," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 5720–5724.
[20] 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.
[21] G. Leus and A.-J. van der Veen, "Optimal training for ML and LMMSE channel estimation in MIMO systems," in *Proc. IEEE/SP 13th Workshop Statist. Signal Process.*, 2005, pp. 1354–1357.
[22] 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, Mar. 2010.
[23] C. Y. Chen and P. P. Vaidyanathan, "MIMO radar waveform optimization with prior information of the extended target and clutter," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3533–3544, Sep. 2009.
[24] O. Aldayel, V. Monga, and M. Rangaswamy, "SQR: Successive QCQP refinement for MIMO radar waveform design under practical constraints," in *Proc. 49th Asilomar Conf. Signals Syst. Comput.*, 2015, pp. 85–89.
[25] P. Stoica, J. Li, and X. Zhu, "Waveform synthesis for diversity-based transmit beampattern design," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2593–2598, Sep. 2008.
[26] J. Liang, L. Xu, J. Li, and P. Stoica, "On designing the transmission and reception of multistatic continuous active sonar systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 285–299, Jan. 2014.
[27] J. Liang, H. C. So, C. S. Leung, J. Li, and A. Farina, "Waveform design with unit modulus and spectral shape constraints via Lagrange programming neural network," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1377–1386, Dec. 2015.
[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993, vol. 2.
[29] 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.
[30] J. H. Manton, "Optimal training sequences and pilot tones for OFDM systems," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 151–153, Apr. 2001.
[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.
[32] M. Golay, "A class of finite binary sequences with alternate auto-correlation values equal to zero (corresp.)," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 449–450, May 1972.
[33] N. Zhang and S. Golomb, "Polyphase sequence with low autocorrelations," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 1085–1089, May 1993.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[35] 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.
[36] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.
[37] 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.
[38] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[39] 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.
[40] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 1995.

Zhongju Wang received the B.Eng. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2013. He is currently working toward the Ph.D. degree in electronic and computer engineering at Hong Kong University of Science and Technology, Hong Kong SAR. His research interests include applications of convex optimization to signal processing and machine learning.

Prabhu Babu received the Ph.D. degree in electrical engineering from the Uppsala University, Sweden, in 2012. He is currently with CARE, IIT Delhi, Delhi, India.

Daniel P. Palomar (S'99–M'03–SM'08–F'12) received the electrical engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively. He is a Professor in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He had previously held several research appointments, namely, at King's College London, London, UK; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia, Barcelona, Spain; Royal Institute of Technology, Stockholm, Sweden; University of Rome La Sapienza, Rome, Italy; and Princeton University, Princeton, NJ, USA. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems, big data systems, and communication systems. Since 2013, he is a Fellow of the Institute for Advance Study (IAS) at HKUST. He received a 2004/06 Fulbright Research Fellowship, the 2004 and 2015 (co-author) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015–16 HKUST Excellence Research Award, the 2002/03 best Ph.D. prize in Information Technologies and Communications by the UPC, the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT. He is a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 2016 Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading and has been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on Convex Optimization for Signal Processing, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on Game Theory in Communication Systems, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on Optimization of MIMO Transceivers for Realistic Communication Networks.