

Effective Low-Complexity Optimization Methods for Joint Phase Noise and Channel Estimation in OFDM

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

Abstract—Phase noise correction is crucial to exploit full advantage of orthogonal frequency-division multiplexing (OFDM) in modern high-data-rate communications. OFDM channel estimation with simultaneous phase noise compensation has therefore drawn much attention and stimulated continuing efforts. Existing methods, however, either have not taken into account the fundamental properties of phase noise or are only able to provide estimates of limited applicability owing to considerable computational complexity. In this paper, we have reformulated the joint phase noise and channel estimation problem in the time domain as opposed to existing frequency-domain approaches, which enables us to develop much more efficient algorithms using the majorization-minimization technique. In addition, we propose two methods based on dimensionality reduction and regularization, respectively, that can adapt to various phase noise levels and signal-to-noise ratio and achieve much lower estimation errors than the benchmarks without incurring much additional computational cost. Several numerical examples with phase noise generated by free-running oscillators or phase-locked loops demonstrate that our proposed algorithms outperform existing methods with respect to both computational efficiency and mean squared error within a large range of SNRs.

Index Terms—Carrier frequency offset (CFO), channel estimation, majorization-minimization (MM), orthogonal frequency division multiplexing (OFDM), phase noise.

I. INTRODUCTION

PROMINENT advantages such as higher spectral efficiency, adaptability to severe channel environments, and efficient implementation have brought orthogonal frequency division multiplexing (OFDM) into wide applications in modern communications. To fully exploit these advantages in reality, we have to resolve some demanding issues—sensitivity to frequency synchronization errors, high peak-to-average power ratios, to name a few. In this paper, we will focus on the frequency synchronization issue stemming specifically from phase noise.

Phase noise is a random process caused by the fluctuation in phase within receiver and transmitter oscillators that are

deployed to generate carrier signals for up-down conversion [1]–[7]. In practice, free-running oscillators and phase-locked loops are widely used, for which phase noise is described by Wiener process and Gaussian process (or Ornstein-Uhlenbeck model in [8]), respectively [8]–[10]. An OFDM block, consisting of several symbols, is transmitted and received with orthogonal subcarriers. Due to the introduction of phase noise, however, the orthogonality among subcarriers is lost, which causes degradation in performance of OFDM systems. Such loss of performance has been well-documented and studied with detailed analyses of, e.g., signal-to-noise ratio (SNR) and bit error rate (BER); see [1]–[3], [8]–[13] and many references therein. For instance, it has been reported in [13] that for OFDM with 2048 subcarriers, Wiener phase noise of 3dB bandwidth $\Delta f_{3\text{dB}}$ of 100 Hz can bring about an SNR degradation over 10 dB, which implies the acute sensitivity of OFDM systems to the existence of phase noise.

Common phase error (CPE) and inter-carrier interference (ICI) are two detrimental effects caused by phase noise. CPE causes phase rotation to each subcarrier and does not change within a transmitted OFDM block. In contrast, ICI introduces different interference to different subcarriers in the same block, and thus exhibits noise-like characteristics [14]. Even with small phase noise, where ICI dominates over CPE, CPE-only correction can provide a 5 dB gain in SNR [3]. To be general, in this paper we will consider phase noise that is not necessarily small. Also, even though a constant carrier frequency offset (CFO) may exist apart from phase noise, we assume CFO has been fixed using readily available methods, e.g., [15].

Many works, e.g., [8], [16], [17], have studied phase noise estimation assuming channel information is given or in the context of transmit data detection. Phase noise estimation and compensation can reduce ICI between data subcarriers and directly facilitate information decoding. In practice, however, phase noise exists throughout the channel training and data transmission stages, and the effects of phase noise to channel estimates also influence the next data detection. Joint estimation of phase noise and channel has thus been proposed to improve the quality of channel estimates [18]–[20]. The joint estimation problem has been investigated as early as in [13]. The least-squares estimator of channel impulse response is computed first; then heuristically, a window function as a filter is applied to the obtained channel estimator to reduce its sensitivity from phase noise and CFO. To be statistically justified, maximum *a posteriori* channel estimator in [18] has exploited the statistical properties of phase noise. But the authors use a Taylor expansion to approximate the

Manuscript received October 5, 2016; revised January 30, 2017; accepted March 10, 2017. Date of publication April 6, 2017; date of current version April 24, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rui Zhang. This work was supported by the Hong Kong RGC 16206315 research grant. (*Corresponding author: Zhongju Wang.*)

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 is with Center for Applied Research in Electronics, Indian Institute of Technology Delhi, New 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.2017.2691672

nonlinear optimization objective function, which works only for small phase noise. A simple alternating optimization method for the joint estimation problem can be found in [19]. The critical issue with that method is its failure to deal with the constraint of phase noise in each iterative sub-problem.

Supposedly, estimating phase noise and channel was hard to disentangle as previous works claimed. In [21], a novel formulation is proposed with phase noise and channel estimations unraveled. To solve the resulting optimization problem, the authors replace the unimodular constraint on phase noise in the time domain with a relaxation assuming the magnitude of phase noise is relatively small. Nevertheless, their method is computationally unstable with a singularity issue that renders the already approximated solution even more inaccurate. And recently, a method craftily using the spectral property of phase noise is provided for the frequency domain-formulated problem [22]. Based on [21], the separate phase noise estimation problem is solved by semidefinite programming (SDP). This method works fine when the number of subcarriers deployed in OFDM is not too large and phase noise arises in a small level. In reality, however, the number of subcarriers can be as large as tens of thousands, e.g., in terrestrial television broadcasting system (DVB-T2) [23].

Regarding the joint phase noise and channel estimation, there are basically two classes of approaches: time-domain [13], [18]–[20], [24]–[26] and frequency-domain approaches [8], [21], [22]. In this paper, we formulate the optimization problem in the time-domain representation and our contributions are as follows. First, we prove the equivalence of the frequency-domain approach and the time-domain approach to the problem formulation. It allows us to separate the joint estimation problem and to focus on estimating phase noise. And using the majorization-minimization technique, we devise more efficient algorithms as opposed to solving an SDP as in [22]. The efficiency and low complexity of our proposed algorithms enable us to deal readily with much larger number of subcarriers. Moreover, we offer two adaptive methods for further reducing estimation errors considering that the joint estimation problem is underdetermined per se irrespective of the approach of formulation. To achieve this, dimensionality reduction has been adopted in [19], [22] to address either the underdetermined nature of the problem or the computational complexity. But instead of adopting a fixed dimension, we run our algorithms with different reduced sizes and opt for the solution that yields the minimal Bayesian information criterion (BIC). Besides dimensionality reduction, we also propose a regularization to the original estimation objective to prevent potential over-fitting. The extra adaptability to various levels of phase noise and SNR comes without incurring much computational burden as simulated examples demonstrate.

The structure of this paper is as follows. We give the system model of OFDM with a description of phase noise in Section II. In Section III, the problem formulation is presented after a review of existing methods. We dedicate Section IV to developing algorithms solving the formulated problem with extensions based on dimensionality reduction and regularization. Simulation results are given in Section V, followed by a conclusion to summarize the paper in Section VI.

We use the following notation throughout this paper. Scalars, vectors, and matrices are denoted by italic letters, boldface lower-case letters, and boldface upper-case letters, respectively. The superscript $(\cdot)^T$ denotes the transpose, $(\cdot)^H$ the conjugate transpose. The ℓ_2 -norm and ℓ_∞ -norm of a vector is denoted by $\|\cdot\|$ and $\|\cdot\|_\infty$, respectively. The identity matrix is denoted by \mathbf{I}_n with size specified by the subscript n . \mathbb{R} is the set of real numbers. $\mathbf{1}_n$ is an all-one vector of length n . $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of a matrix. $\mathcal{O}(\cdot)$ denotes Big-O notation.

II. SYSTEM MODEL AND DESCRIPTION OF PHASE NOISE

A. OFDM Transmission Model

Suppose there are N_c subcarriers and an OFDM block is denoted by $\mathbf{s} = [s_0, \dots, s_{N_c-1}]^T$. The time-domain symbols can be obtained by the unitary inverse discrete Fourier transform (IDFT):

$$x_n = \frac{1}{\sqrt{N_c}} \sum_{k=0}^{N_c-1} s_k e^{j\frac{2\pi nk}{N_c}}, \quad n = 0, 1, \dots, N_c - 1. \quad (1)$$

Let \mathbf{F} be the $N_c \times N_c$ unitary discrete Fourier transform (DFT) matrix, then (1) can be written as

$$\mathbf{x} = \mathbf{F}^H \mathbf{s}. \quad (2)$$

Assume a slow-varying channel whose response does not change within the transmission of several OFDM blocks and is denoted by $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}]^T$ ($N_c \gg L$). Let $\mathbf{x} = [x_0, x_1, \dots, x_{N_c-1}]^T$ and $\mathbf{y} = [y_0, y_1, \dots, y_{N_c-1}]^T$. With the cyclic prefix appending and removal, we have the OFDM transmission model [27, Ch. 3.4.4]:

$$\mathbf{y} = \mathbf{x} \circledast \begin{bmatrix} \mathbf{h} \\ \mathbf{0} \end{bmatrix} + \mathbf{v}, \quad (3)$$

where \circledast denotes the operation of circular convolution, and $\mathbf{v} = [v_0, v_1, \dots, v_{N_c-1}]^T$ is a zero-mean circularly symmetric complex Gaussian channel noise vector with distribution $\mathcal{CN}(\mathbf{0}, 2\sigma^2 \mathbf{I})$.

To obtain the frequency-domain representation of (3), take the DFT to both sides and we have¹

$$\mathbf{r} = \sqrt{N_c} \mathbf{H} \mathbf{s} + \mathbf{w}, \quad (4)$$

where \mathbf{r} is the unitary DFT of the received time-domain symbols \mathbf{y} , \mathbf{H} is a diagonal matrix with the N_c -point unitary DFT of \mathbf{h} as the diagonal, and \mathbf{w} is the unitary DFT of the time-domain channel noise \mathbf{v} . Let $\tilde{\mathbf{F}}$ be a semi-unitary matrix formed by the first L columns of \mathbf{F} , then $\mathbf{H} = \text{Diag}(\tilde{\mathbf{F}}\mathbf{h})$.

B. OFDM Transmission With Phase Noise

In general, phase noise is present in the local oscillators that generate carrier signals for up-down conversion for the time-domain symbols. And the effect of phase noise can be represented mathematically by multiplying each time-domain symbol with a complex exponential with a random phase. Although phase noise exists in both the transmitter and the receiver, herein

¹Note that the factor $\sqrt{N_c}$ results from using the unitary DFT.

only the effect at the receiver side is studied. The reason for this simplified consideration is the assumption that at the transmitter side, the bandwidth of phase noise is small [14] or high-caliber oscillators are employed [22]. Therefore, the following signal model with phase noise is considered [8]:

$$\mathbf{y} = e^{j\boldsymbol{\theta}} \odot \left(\mathbf{x} \otimes \begin{bmatrix} \mathbf{h} \\ \mathbf{0} \end{bmatrix} \right) + \mathbf{v}, \quad (5)$$

where $e^{j\boldsymbol{\theta}} := [e^{j\theta_0}, e^{j\theta_1}, \dots, e^{j\theta_{N_c-1}}]^T$ denotes phase noise, and \odot denotes the Hadamard product. Taking the unitary DFT on both sides of (5), we can obtain the frequency-domain signal model:

$$\mathbf{r} = \boldsymbol{\phi} \otimes (\mathbf{H}\mathbf{s}) + \mathbf{w}, \quad (6)$$

where $\boldsymbol{\phi} = [\phi_0, \phi_1, \dots, \phi_{N_c-1}]^T = \mathbf{F}e^{j\boldsymbol{\theta}}$, called spectral phase noise vector, and \mathbf{w} are the unitary DFT of $e^{j\boldsymbol{\theta}}$ and \mathbf{v} , respectively. For each received frequency-domain symbol r_k , $k = 0, 1, \dots, N_c - 1$, we have

$$r_k = \phi_0 H_{k,k} s_k + \sum_{l=0, l \neq k}^{N_c-1} \phi_{k-l} H_{l,l} s_l + w_k, \quad (7)$$

where the first term, subjected to the scaling of factor ϕ_0 , is called CPE, and the second term, combining effects from other subcarriers, is ICI. With $r_k = r_{k \bmod N_c}$, (6) can be rewritten in the following matrix form

$$\mathbf{r} = \boldsymbol{\Phi}_\phi \mathbf{H}\mathbf{s} + \mathbf{w}, \quad (8)$$

in which

$$\boldsymbol{\Phi}_\phi = \begin{bmatrix} \phi_0 & \phi_{N_c-1} & \cdots & \phi_2 & \phi_1 \\ \phi_1 & \phi_0 & \cdots & \phi_3 & \phi_2 \\ \vdots & & \ddots & & \vdots \\ \phi_{N_c-2} & \phi_{N_c-3} & \cdots & \phi_0 & \phi_{N_c-1} \\ \phi_{N_c-1} & \phi_{N_c-2} & \cdots & \phi_1 & \phi_0 \end{bmatrix}, \quad (9)$$

denoted by $\boldsymbol{\Phi}_\phi = \text{circ}(\boldsymbol{\phi})$, is a circulant matrix formed by spectral phase noise $\boldsymbol{\phi}$. The off-diagonals of $\boldsymbol{\Phi}_\phi$ close to the main diagonal correspond to low-frequency components. With $\mathbf{H} = \text{Diag}(\check{\mathbf{F}}\mathbf{h})$, (8) can be rewritten as

$$\mathbf{r} = \boldsymbol{\Phi}_\phi \mathbf{S}\check{\mathbf{F}}\mathbf{h} + \mathbf{w}, \quad (10)$$

where $\mathbf{S} = \text{Diag}(\mathbf{s})$ is a diagonal matrix with \mathbf{s} as the diagonal and $\check{\mathbf{F}}$ a semi-unitary matrix with the first L columns of \mathbf{F} .

C. Properties of Phase Noise

Two canonical models of phase noise are Wiener process and Gaussian process when free-running oscillators and phase-locked loops are respectively employed [8]. The statistical properties of phase noise have also been studied in [8], [14]. Before introducing some existing formulations of the phase noise estimation problem, some useful properties of phase noise are presented here.

1) *Time-Domain Property*: Obviously, phase noise $e^{j\boldsymbol{\theta}}$ is determined only by the phase variable $\boldsymbol{\theta}$, and phase noise at each OFDM subcarrier is *unimodular*, i.e.,

$$|e^{j\theta_n}| = 1, \quad n = 0, 1, \dots, N_c - 1. \quad (11)$$

2) *Frequency-Domain Property*: Let $\boldsymbol{\phi}$ and $\underline{\boldsymbol{\phi}}$ be the unitary DFT of $e^{j\boldsymbol{\theta}}$ and $e^{-j\boldsymbol{\theta}}$, respectively. It is well-known that $\boldsymbol{\phi}$ and $\underline{\boldsymbol{\phi}}$ are conjugate symmetric, i.e.,

$$\underline{\phi}_k = \phi_{-k}^*. \quad (12)$$

Observing that $e^{j\boldsymbol{\theta}} \odot e^{-j\boldsymbol{\theta}} = \mathbf{1}_{N_c}$ and applying the DFT to both sides, we can obtain the following constraint for spectral phase noise:

$$\boldsymbol{\phi} \otimes \underline{\boldsymbol{\phi}} = N_c \boldsymbol{\delta}_k, \quad (13)$$

where δ_k is the Kronecker delta function, i.e., $\delta_0 = 1$, and $\delta_k = 0$ for $k \neq 0$. Indeed, (13) is a necessary and sufficient description of the autocorrelation of the spectral components of any unimodular complex exponential sequence, which can be easily verified by Fourier transform and its properties. Equivalently, (13) can be written in a matrix form as

$$\boldsymbol{\Phi}_\phi^H \boldsymbol{\Phi}_\phi = N_c \mathbf{I}_{N_c}, \quad (14)$$

where $\boldsymbol{\Phi}_\phi = \text{circ}(\boldsymbol{\phi})$ is a circulant matrix defined in (9). This is the main property exploited in [22], termed the spectral geometry.

III. LITERATURE REVIEW AND PROBLEM FORMULATION

Phase noise contamination can be removed from the received OFDM symbols if a reliable estimate of the instantaneous realization of phase noise process is accessible. In practice, phase noise exists throughout channel estimation stage and data detection stage, where phase noise estimation is entangled with the unknown channel and unknown transmitted data, respectively. Plenty of works are available for estimating transmitted data with phase noise compensated, e.g., [19], [24]. In this paper we will focus on joint phase noise and channel estimation, where transmitted symbols (called pilot symbols or training symbols) are assumed known to receiver [27, Ch. 3.5.2]. A motivation for studying this joint estimation problem is that assuming channel is quasi-static or slowly-varying, the channel estimate can be reasonably used in the subsequent data estimation [19]. In the literature, methods for this purpose can be categorized into two classes: time-domain approach and frequency-domain approach. In particular, we assume in this paper that phase noise $\boldsymbol{\theta}$ and channel impulse response \mathbf{h} are constant parameters to estimate, and OFDM symbols \mathbf{S} are given and known to the receiver.

A. Time-Domain Approaches

In [19], the authors formulate the least-squares problem with (10)

$$\underset{\mathbf{h}, \boldsymbol{\theta}, \boldsymbol{\phi} = \mathbf{F}e^{j\boldsymbol{\theta}}}{\text{minimize}} \left\| \mathbf{r} - \boldsymbol{\Phi}_\phi \mathbf{S}\check{\mathbf{F}}\mathbf{h} \right\|^2, \quad (15)$$

and solve for channel and phase noise estimates alternately. At the i th iteration, given the phase noise estimate $e^{j\hat{\theta}^{(i-1)}}$, the channel estimate is computed by

$$\hat{\mathbf{h}}^{(i)} = \left(\check{\mathbf{F}}^H \mathbf{S}^H (\Phi_{\phi}^{(i-1)})^H \Phi_{\phi}^{(i-1)} \mathbf{S} \check{\mathbf{F}} \right)^{-1} \check{\mathbf{F}}^H \mathbf{S}^H (\Phi_{\phi}^{(i-1)})^H \mathbf{r} \quad (16)$$

with $\Phi_{\phi}^{(i-1)} = \text{circ}(\mathbf{F}e^{j\hat{\theta}^{(i-1)}})$. Let $\mathbf{c} = e^{j\theta}$, then the estimate for phase noise is updated as

$$\hat{\mathbf{c}}^{(i)} = (\mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F})^{-1} \mathbf{F}^H \mathbf{P}^H \mathbf{r}, \quad (17)$$

where $\mathbf{P} = \text{circ}(\mathbf{S}\check{\mathbf{F}}\hat{\mathbf{h}}^{(i)})$. Yet there are two issues with their method: the unimodular property of phase noise vector is not considered when updating $\hat{\mathbf{c}}^{(i)}$; and the alternating optimization scheme suffers from slow convergence.

Some other heuristic methods include approximating phase noise by a Taylor expansion [18], applying filtering to channel estimate with a noise-suppressing function [13], approximating with sinusoidal waveforms [24], and Monte Carlo methods [20], [25], [26].

B. Frequency-Domain Approaches

In [8], a phase noise correction method is proposed by estimating the spectral components, based on the assumption that phase noise process can be characterized by a low-pass signal and thus only a few spectral components need to be estimated. But it is necessary to find a proper number of spectral phase noise components in order to achieve reliable estimation. Although [8] also exploits the statistical properties of ICI to obtain the MMSE estimate of phase noise, their method is subject to two main issues: the channel is assumed known and the MMSE estimation has not taken into account the constraint (13) of spectral phase noise.

Following the same idea of [8] to estimate the low-frequency components of phase noise, [21] formulates the problem of joint phase noise and channel estimation based on least-squares. To acquire separate estimators, instead of alternately updating (16) and (17), they substitute the channel estimate into the least-squares objective and the resulting error function for phase noise can be derived as

$$\mathcal{E}(\phi) = \mathbf{r}^H \mathbf{r} - \frac{1}{N_c} \mathbf{r}^H \Phi_{\phi} \mathbf{B} \Phi_{\phi}^H \mathbf{r} \quad (18)$$

$$= \frac{1}{N_c} \phi^H \mathbf{J}_1 (\mathbf{R}^H \mathbf{R} - \mathbf{R}^H \mathbf{B} \mathbf{R})^T \mathbf{J}_1 \phi, \quad (19)$$

where $\mathbf{B} = \check{\mathbf{F}} (\check{\mathbf{F}}^H \mathbf{S}^H \check{\mathbf{F}})^{-1} \check{\mathbf{F}}^H \mathbf{S}^H$ and \mathbf{J}_1 a permutation matrix, left multiplication by which keeps the first row and reverses the orders of the remaining rows. Note that in [21], the expression for $\mathcal{E}(\phi)$ is further simplified assuming the transmitted symbols are of constant-modulus. When solving for the phase noise estimate, however, an approximation by a Taylor expansion is applied, which leads to a relaxed constraint on phase noise. In practice, this approximation works only for small phase noise.

In contrast, [22] incorporates the fundamental spectral constraint (14) into the formulation proposed in [21]. Let

$$\mathbf{M} = \frac{1}{N_c} \mathbf{J}_1 (\mathbf{R}^H \mathbf{R} - \mathbf{R}^H \mathbf{B} \mathbf{R})^T \mathbf{J}_1, \quad (20)$$

then the problem is formulated as

$$\begin{aligned} & \underset{\phi}{\text{minimize}} && \phi^H \mathbf{M} \phi \\ & \text{subject to} && \Phi_{\phi}^H \Phi_{\phi} = N_c \mathbf{I}_{N_c}, \Phi_{\phi} = \text{circ}(\phi). \end{aligned} \quad (21)$$

Instead of solving (21), dimensionality reduction is introduced to alleviate the computation complexity by estimating only the low-frequency components, cf. [8]. To achieve this, the phase-noise-geometry preserving transformation is defined by

$$\phi = \mathbf{T} \check{\phi}, \quad (22)$$

where $\check{\phi}$ of a shorter length N is the reduced spectral phase noise to be estimated. An example of \mathbf{T} is piecewise-constant transformation (PCT). Then an alternative optimization problem is posed as follows:

$$\begin{aligned} & \underset{\check{\phi}}{\text{minimize}} && \check{\phi}^H \mathbf{T}^H \mathbf{M} \mathbf{T} \check{\phi} \\ & \text{subject to} && \check{\Phi}_{\check{\phi}}^H \check{\Phi}_{\check{\phi}} = N \mathbf{I}_N, \check{\Phi}_{\check{\phi}} = \text{circ}(\check{\phi}). \end{aligned} \quad (23)$$

To solve the above problem, the S-procedure is invoked to rewrite (23) as a semidefinite program (SDP). The original spectral phase noise vector ϕ can be recovered by (22). To guarantee the constraint (14) still holds, the authors provide a sufficient condition for the transformation matrix. Their method, however, suffers from several limitations. When the reduced length N is not small enough, SDP reformulation still renders a solution failing to satisfy the spectral constraint of phase noise; yet, it is prohibited to solve a large dimensional SDP. Nowadays, the number of subcarriers can be up to thousands and to use this method, the original dimension needs to be greatly reduced, which can result in the loss of reliability and accuracy in the obtained estimate. Furthermore, the reduced spectral phase noise does not necessarily satisfy the spectral constraint as imposed in problem (23); thus, this method gives a tightened solution.

C. Problem Formulation

Based on (5), the time-domain OFDM model with phase noise is given by

$$\mathbf{y} = \sqrt{N_c} \text{Diag}(e^{j\theta}) \mathbf{F}^H \mathbf{S} \check{\mathbf{F}} \mathbf{h} + \mathbf{v}. \quad (24)$$

Similar to (15), we propose the following optimization problem:

$$\underset{\mathbf{h}, \theta}{\text{minimize}} \quad \|\mathbf{y} - \sqrt{N_c} \text{Diag}(e^{j\theta}) \mathbf{F}^H \mathbf{S} \check{\mathbf{F}} \mathbf{h}\|^2. \quad (25)$$

Solving (25) for \mathbf{h} gives the least-squares channel estimate

$$\hat{\mathbf{h}} = \frac{1}{\sqrt{N_c}} (\check{\mathbf{F}}^H \mathbf{S}^H \check{\mathbf{F}})^{-1} \check{\mathbf{F}}^H \mathbf{S}^H \mathbf{F} \text{Diag}(e^{j\theta})^H \mathbf{y}. \quad (26)$$

And the resulting least-squares error for phase noise is

$$\mathcal{E}(\theta) = \mathbf{y}^H \text{Diag}(e^{j\theta}) \mathbf{F}^H (\mathbf{I}_{N_c} - \mathbf{B}) \mathbf{F} \text{Diag}(e^{j\theta})^H \mathbf{y}, \quad (27)$$

where $\mathbf{B} = \mathbf{S}\check{\mathbf{F}}(\check{\mathbf{F}}^H\mathbf{S}^H\mathbf{S}\check{\mathbf{F}})^{-1}\check{\mathbf{F}}^H\mathbf{S}^H$. The phase noise estimation problem is thus formulated as

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \mathbf{y}^H \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H (\mathbf{I}_{N_c} - \mathbf{B}) \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}})^H \mathbf{y}. \quad (28)$$

Let us introduce $\mathbf{V} = \mathbf{F}^H (\mathbf{I}_{N_c} - \mathbf{B}) \mathbf{F}$ and $\mathbf{u} = e^{-j\boldsymbol{\theta}}$. We can rewrite (28) as the following quadratic problem:

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} \quad \mathbf{u}^H \text{Diag}(\mathbf{y})^H \mathbf{V} \text{Diag}(\mathbf{y}) \mathbf{u} \\ & \text{subject to} \quad |u_n| = 1, \quad n = 0, 1, \dots, N_c - 1. \end{aligned} \quad (29)$$

Consequently, the joint phase noise and channel estimation problem boils down to the phase noise estimation problem (29) followed by computing the channel estimate with (26).

D. Equivalence of Time- and Frequency-Domain Approaches

In this section, we show that our formulation of the joint estimation problem (25) and the resulting phase noise estimation problem (29) are equivalent to the existing approaches.

Lemma 1: Let $\boldsymbol{\phi} = \mathbf{F}e^{j\boldsymbol{\theta}}$ and $\Phi_{\boldsymbol{\phi}} = \text{circ}(\boldsymbol{\phi})$, then $\Phi_{\boldsymbol{\phi}} = \sqrt{N_c} \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H$.

Proof: According to the eigenvalue decomposition of a circulant matrix [28],

$$\begin{aligned} \Phi_{\boldsymbol{\phi}} &= \mathbf{F} \text{Diag}(\sqrt{N_c} \mathbf{F} (\mathbf{e}_1^T \Phi_{\boldsymbol{\phi}})^T) \mathbf{F}^H \\ &= \mathbf{F} \text{Diag}(\sqrt{N_c} \mathbf{F}^H (\mathbf{e}_1^T \Phi_{\boldsymbol{\phi}})^H)^* \mathbf{F}^H \\ &= \mathbf{F} \text{Diag}(\sqrt{N_c} \mathbf{F}^H \boldsymbol{\phi})^* \mathbf{F}^H \\ &= \mathbf{F} \text{Diag}(\sqrt{N_c} e^{-j\boldsymbol{\theta}})^* \mathbf{F}^H \\ &= \sqrt{N_c} \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H, \end{aligned}$$

where $\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]^T$. \blacksquare

With Lemma 1, we can prove that the objective function in problem (25) is the same as that of (15):

$$\left\| \mathbf{y} - \sqrt{N_c} \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H \mathbf{S}\check{\mathbf{F}}\mathbf{h} \right\|^2 = \left\| \mathbf{F}^H \mathbf{r} - \sqrt{N_c} \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H \mathbf{S}\check{\mathbf{F}}\mathbf{h} \right\|^2 \quad (30)$$

$$= \left\| \mathbf{r} - \sqrt{N_c} \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H \mathbf{S}\check{\mathbf{F}}\mathbf{h} \right\|^2 \quad (31)$$

$$= \left\| \mathbf{r} - \Phi_{\boldsymbol{\phi}} \mathbf{S}\check{\mathbf{F}}\mathbf{h} \right\|^2. \quad (32)$$

Since

$$\Phi_{\boldsymbol{\phi}}^H \mathbf{r} = \sqrt{N_c} \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}})^H \mathbf{F}^H \mathbf{r} = \sqrt{N_c} \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}})^H \mathbf{y}, \quad (33)$$

(27) is equivalent to the frequency-domain phase noise error function (18):

$$\mathcal{E}(\boldsymbol{\phi}) = \frac{1}{N_c} \mathbf{r}^H \Phi_{\boldsymbol{\phi}} (\mathbf{I}_{N_c} - \mathbf{B}) \Phi_{\boldsymbol{\phi}}^H \mathbf{r} \quad (34)$$

$$= \mathbf{y}^H \text{Diag}(e^{j\boldsymbol{\theta}}) \mathbf{F}^H (\mathbf{I}_{N_c} - \mathbf{B}) \mathbf{F} \text{Diag}(e^{j\boldsymbol{\theta}})^H \mathbf{y}. \quad (35)$$

In the next section, we will use the majorization-minimization technique to develop efficient algorithms to solve problem (29).

IV. ALGORITHMS

A. The Majorization-Minimization Technique

The majorization-minimization (MM) technique provides an approximation-based iterative approach to solving an optimization problem of a generic form [29]–[31]. As the original problem is difficult to address directly, the MM technique follows an iterative procedure—a simpler surrogate objective function is minimized in each iteration—to find a local optimum.

Consider the problem of

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in \mathcal{X}. \quad (36)$$

The MM technique starts from a feasible point $\mathbf{x}^{(0)} \in \mathcal{X}$, and solves a series of simpler majorized problems:

$$\underset{\mathbf{x}}{\text{minimize}} \quad g(\mathbf{x}; \mathbf{x}^{(t)}) \quad \text{subject to} \quad \mathbf{x} \in \mathcal{X}, \quad (37)$$

$t = 0, 1, \dots$, each of which produces an updated point $\mathbf{x}^{(t+1)}$. Basically, the surrogate objective, known as the majorization function for $f(\mathbf{x})$, should satisfy the following conditions:

$$g(\mathbf{x}^{(t)}; \mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \quad (38)$$

$$g(\mathbf{x}; \mathbf{x}^{(t)}) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}, \quad (39)$$

$$\nabla_{\mathbf{d}} g(\mathbf{x}^{(t)}; \mathbf{x}^{(t)}) = \nabla_{\mathbf{d}} f(\mathbf{x}^{(t)}) \quad \forall \mathbf{x}^{(t)} + \mathbf{d} \in \mathcal{X}, \quad (40)$$

where $\nabla_{\mathbf{d}} g(\mathbf{x}^{(t)}; \mathbf{x}^{(t)})$ is the directional derivative of g at $\mathbf{x}^{(t)}$ in the direction of \mathbf{d} . Consequently, a series of points that result in non-increasing objective values are obtained:

$$f(\mathbf{x}^{(t+1)}) \leq g(\mathbf{x}^{(t+1)}; \mathbf{x}^{(t)}) \leq g(\mathbf{x}^{(t)}; \mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \quad (41)$$

And any limit point of thus generated sequence of points is a stationary solution to the original problem (36).

To develop an efficient MM-based algorithm, the series of problems (37) should all be simple enough to solve—ideally, each should be solved with a closed-form solution. Crucial to achieve such a goal is to find a good majorization function $g(\mathbf{x}; \mathbf{x}^{(t)})$, which requires to properly exploit the particular structure of the specific problem. Some general and useful rules for majorization can be found in [31]. In the next section, we will devise MM algorithms to solve our problem (29) with two different majorizing methods. Also it will be illustrated in simulations that the majorization is critical for the convergence speed of the obtained algorithms.

B. The MM Algorithms for Phase Noise Estimation

The following lemma is introduced first, which is useful for finding majorization functions.

Lemma 2: Given a matrix \mathbf{A} , $\mathbf{P} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ is an orthogonal projection matrix, which is unitarily similar to a diagonal matrix with diagonal entries being either 1 or 0 [32, Corollary 3.4.3.3].

1) *Loose Quadratic Majorization (LQM)*: Let us write $\tilde{\mathbf{V}} = \text{Diag}(\mathbf{y})^H \mathbf{V} \text{Diag}(\mathbf{y})$. The objective in (29) can be majorized by a quadratic function at \mathbf{u}_0 as follows [33, Lemma 1]:

$$\mathbf{u}^H \tilde{\mathbf{V}} \mathbf{u} \leq 2\text{Re} \left\{ \mathbf{u}_0^H \left(\tilde{\mathbf{V}} - \lambda \mathbf{I}_{N_c} \right) \mathbf{u} \right\} + 2\lambda \|\mathbf{u}\|^2 - \mathbf{u}_0^H \tilde{\mathbf{V}} \mathbf{u}_0, \quad (42)$$

in which $\lambda \mathbf{I}_{N_c} \succeq \tilde{\mathbf{V}}$ for some constant λ . Note that the largest eigenvalue of \mathbf{V} is 1 by Lemma 2, then we have $\lambda_{\max}(\tilde{\mathbf{V}}) \leq \|\mathbf{y}\|_\infty^2$. Choosing $\lambda = \|\mathbf{y}\|_\infty^2$ will thus satisfy the majorization condition. At the step t , the following majorized problem with the surrogate objective function is solved (since $\|\mathbf{u}\|^2$ is just a constant):

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} && -2\text{Re} \left\{ \left(\mathbf{u}^{(t)} \right)^H \left(\|\mathbf{y}\|_\infty^2 \mathbf{I}_{N_c} - \tilde{\mathbf{V}} \right) \mathbf{u} \right\} \\ & \text{subject to} && |u_n| = 1, \quad n = 0, 1, \dots, N_c - 1. \end{aligned} \quad (43)$$

It is obvious that a closed-form solution to (43) is:

$$\begin{aligned} \mathbf{u}^{(t+1)} &= \exp \left[j \arg \left(\left(\|\mathbf{y}\|_\infty^2 \mathbf{I}_{N_c} - \tilde{\mathbf{V}} \right) \mathbf{u}^{(t)} \right) \right] \\ &= \exp \left[j \arg \left(\left(\|\mathbf{y}\|_\infty^2 \mathbf{1} - |\mathbf{y}|^2 \right) \odot \mathbf{u}^{(t)} \right. \right. \end{aligned} \quad (44)$$

$$\left. \left. + \text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{u}^{(t)} \right) \right], \quad (45)$$

where the exponential and the squared magnitude $|\cdot|^2$ are taken element-wise. We call this method a loose quadratic majorization (LQM) because the structure of the original objective function could have been better exploited as shown below, which leads to faster convergence.

2) *Tight Quadratic Majorization (TQM)*: Similar to (42), the original objective can be majorized as follows:

$$\begin{aligned} & \mathbf{u}^H \text{Diag}(\mathbf{y})^H \mathbf{V} \text{Diag}(\mathbf{y}) \mathbf{u} \\ & \leq \lambda \mathbf{u}^H \text{Diag}(\mathbf{y})^H \text{Diag}(\mathbf{y}) \mathbf{u} \\ & \quad + 2\text{Re} \left\{ \mathbf{u}_0^H \text{Diag}(\mathbf{y})^H \left(\mathbf{V} - \lambda \mathbf{I}_{N_c} \right) \text{Diag}(\mathbf{y}) \mathbf{u} \right\} \\ & \quad + \mathbf{u}_0^H \text{Diag}(\mathbf{y})^H \left(\lambda \mathbf{I}_{N_c} - \mathbf{V} \right) \text{Diag}(\mathbf{y}) \mathbf{u}_0 \end{aligned} \quad (46)$$

$$\begin{aligned} & = 2\text{Re} \left\{ \mathbf{u}_0^H \text{Diag}(\mathbf{y})^H \left(\mathbf{V} - \lambda \mathbf{I}_{N_c} \right) \text{Diag}(\mathbf{y}) \mathbf{u} \right\} \\ & \quad + 2\lambda \|\mathbf{y}\|^2 - \mathbf{u}_0^H \text{Diag}(\mathbf{y})^H \mathbf{V} \text{Diag}(\mathbf{y}) \mathbf{u}_0, \end{aligned} \quad (47)$$

where $\lambda \mathbf{I}_{N_c} \succeq \mathbf{V}$ for some constant λ and the equality follows from the unimodular property of $u_n, n = 0, 1, \dots, N_c - 1$. To find a good majorization function, we can choose $\lambda = 1$ by Lemma 2. At the step t , the following majorized problem can be obtained:

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} && -2\text{Re} \left\{ \left(\mathbf{u}^{(t)} \right)^H \text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{u} \right\} \\ & \text{subject to} && |u_n| = 1, \quad n = 0, 1, \dots, N_c - 1, \end{aligned} \quad (48)$$

which results in a closed-form solution:

$$\mathbf{u}^{(t+1)} = \exp \left[j \arg \left(\text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{u}^{(t)} \right) \right]. \quad (49)$$

It will be demonstrated later that this method converges faster owing to its tighter majorization.

Algorithm 1: Algorithm for Phase Noise Estimation with TQM.

- 1: Compute $\mathbf{B} = \mathbf{S} \tilde{\mathbf{F}} \left(\tilde{\mathbf{F}}^H \mathbf{S}^H \mathbf{S} \tilde{\mathbf{F}} \right)^{-1} \tilde{\mathbf{F}}^H \mathbf{S}^H$, set $t = 0$, and initialize $\mathbf{u}^{(0)} = e^{j\theta_0}$.
 - 2: **repeat**
 - 3: $\mathbf{u}^{(t+1)} = \exp \left[j \arg \left(\text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{u}^{(t)} \right) \right]$
 - 4: $t \leftarrow t + 1$
 - 5: **until** convergence.
-

The whole procedure is summarized in Algorithm 1 for TQM. Since main difference from TQM lies in the update of $\mathbf{u}^{(t+1)}$, the algorithm for LQM is omitted here. Once the algorithm converges to solution \mathbf{u}^* , the phase noise estimate can be obtained by

$$\hat{\theta} = -\arg(\mathbf{u}^*). \quad (50)$$

Remark 1: There exists phase rotation ambiguity in problem (25): it can be seen that phase noise and channel estimates are subjected to reciprocal common phase rotations. Let $\hat{\mathbf{h}}$ and $\hat{\theta}$ be channel and phase noise estimates, respectively. The least-squares error of the estimates $\hat{\mathbf{h}}$ and $\hat{\theta}$ is the same as that of $e^{j\theta_c} \hat{\mathbf{h}}$ and $\hat{\theta} - \theta_c \mathbf{1}_{N_c}$. Since θ_c keeps unchanged among subcarriers, it acts like CFO. Many effective methods can be found for CFO correction; see, e.g., [15], [18]. Assuming CFO has been eliminated before estimating phase noise, we can thus set $\theta_c = 0$. Therefore, once Algorithm 1 converges to a solution \mathbf{u}^* , phase ambiguity can be removed by the rotation: $\mathbf{u}^* \leftarrow \mathbf{u}^* / u_0^*$.

C. Dimensionality Reduction

Dimensionality reduction has been proposed in [22] to alleviate the computational complexity when solving an SDP of size N_c , the number of OFDM subcarriers. More important, as noted in [19], estimation problem (15) and equivalent (25) are essentially underdetermined. To obtain reasonable estimates, the number of unknowns in the problem needs to be reduced and, hence, a reduced phase noise vector is estimated.

Similar to the transformation (22) introduced in [22] for estimating reduced spectral phase noise, we apply dimensionality reduction to our problem (29) in the time domain. Recall $\mathbf{u} = e^{-j\theta}$ for phase noise θ . We define

$$\hat{\mathbf{u}} = \mathbf{T}_N \check{\mathbf{u}} = \mathbf{T}_N e^{-j\check{\theta}} \quad (51)$$

as a mapping from a low-dimensional phase noise $\check{\theta} \in \mathbb{R}^N$ to the original phase noise with the transformation matrix $\mathbf{T}_N \in \mathbb{R}^{N_c \times N}$ ($N < N_c$). Two instances of \mathbf{T}_N are suggested in [22]—piecewise-constant transformation (PCT) and random perturbator. And it has been demonstrated that PCT, albeit simple, achieves the best performance. PCT is defined as

$$\mathbf{T}_N = \begin{bmatrix} \mathbf{1}_{N_s} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{N_s} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{N_s} \end{bmatrix}, \quad (52)$$

with $N_s = N_c/N$. In this case, the transformation matrix acts like a sample-and-hold circuit to recover the desired phase noise. Another transformation matrix is provided in [19] based on linear interpolation. For simplicity, we focus on PCT in this paper, but the methods proposed in the following also apply to the linear interpolation transformation and a numerical example is provided later.

Introducing dimensionality reduction requires us to solve a different optimization problem. By substituting (51) into (29), we can obtain an estimation problem of a lower dimension. A similar procedure, however, can be followed when majorizing the new objective function and developing the MM algorithms. For TQM, the update (49) is modified accordingly as

$$\tilde{\mathbf{u}}^{(t+1)} = e^{j \arg(\mathbf{T}_N^H \text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{T}_N \tilde{\mathbf{u}}^{(t)})}. \quad (53)$$

For LQM, the majorization function needs to be recomputed as the condition $\lambda \mathbf{I}_N \succeq \mathbf{T}_N^H \tilde{\mathbf{V}} \mathbf{T}_N$ involves \mathbf{T}_N . Notice that $\lambda_{\max}(\tilde{\mathbf{V}}) = \|\mathbf{y}\|_\infty^2$, and λ can be set to be $\|\mathbf{y}\|_\infty^2 \lambda_{\max}(\mathbf{T}_N^H \mathbf{T}_N)$ such that the majorization inequality constraint is satisfied. As a result, the update for LQM is obtained as follows:

$$\tilde{\mathbf{u}}^{(t+1)} = e^{j \arg((\|\mathbf{y}\|_\infty^2 \lambda_{\max}(\mathbf{T}_N^H \mathbf{T}_N) \mathbf{I}_N - \mathbf{T}_N^H \tilde{\mathbf{V}} \mathbf{T}_N) \tilde{\mathbf{u}}^{(t)})}. \quad (54)$$

For our chosen PCT, $\lambda_{\max}(\mathbf{T}_N^H \mathbf{T}_N) = N_s$ and (54) simplifies to

$$\tilde{\mathbf{u}}^{(t+1)} = \exp \left[j \arg \left(\left(\|\mathbf{y}\|_\infty^2 N_s \mathbf{I}_N - \mathbf{T}_N^H \tilde{\mathbf{V}} \mathbf{T}_N \right) \tilde{\mathbf{u}}^{(t)} \right) \right]. \quad (55)$$

The transformation matrix requires the reduced dimension N to be specified in advance. Previous works have assumed a fixed PCT with given N , which is hardly flexible to different SNR. Here, we prescribe a set of values of N and run our algorithm for each of those values. In particular, N is chosen as a factor of N_c such that PCT is well-defined. To choose the optimal N , we employ the BIC rule [34], which has been demonstrated very effective in model order selection to avoid over-fitting. For each estimate $\hat{\mathbf{u}}$, the corresponding BIC is defined as

$$\text{BIC}(\hat{\mathbf{u}}) = -2 \ln p(\mathbf{y}, \hat{\mathbf{u}}) + N \ln N_c, \quad (56)$$

where $p(\mathbf{y}, \tilde{\mathbf{u}})$ is the probability density function of \mathbf{y} given $\tilde{\mathbf{u}}$. With model (5) and transformation (51), (56) can be rewritten as

$$\text{BIC}(\hat{\mathbf{u}}) = \frac{\mathcal{E}(\hat{\boldsymbol{\theta}})}{\sigma^2} + N \ln N_c, \quad (57)$$

where $\mathcal{E}(\hat{\boldsymbol{\theta}})$ is the least-squares error (27) of the phase noise estimate $\hat{\mathbf{u}}$. The optimal PCT is then defined as the one that produces the minimal BIC. In doing so, improved estimates are expected, compared with the traditional methods [19], [22]. Furthermore, the computational efficiency of LQM and TQM also guarantees an acceptable computational cost. The whole procedure is described in Algorithm 2.

D. Regularization

Apart from dimensionality reduction, another approach to addressing the potential over-fitting inherent in (25) is to impose

Algorithm 2: Phase Noise Estimation with TQM and the Optimal PCT Selected by BIC.

- 1: Compute $\mathbf{B} = \mathbf{S} \tilde{\mathbf{F}} (\tilde{\mathbf{F}}^H \mathbf{S}^H \tilde{\mathbf{S}} \tilde{\mathbf{F}})^{-1} \tilde{\mathbf{F}}^H \mathbf{S}^H$. Choose \mathbb{T} as a set of PCT matrices of different reduced length N .
 - 2: **for** each $\mathbf{T}_N \in \mathbb{T}$ **do**
 - 3: set $t = 0$ and initialize $\tilde{\mathbf{u}}^{(0)}$
 - 4: **repeat**
 - 5: $\tilde{\mathbf{u}}^{(t+1)} = e^{j \arg(\mathbf{T}_N^H \text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{T}_N \tilde{\mathbf{u}}^{(t)})}$
 - 6: $t \leftarrow t + 1$
 - 7: **until** convergence
 - 8: $\hat{\mathbf{u}} = \mathbf{T}_N \tilde{\mathbf{u}}^{(t+1)}$
 - 9: choose $\hat{\mathbf{u}}$ with the minimal BIC ($\hat{\mathbf{u}}$).
-

a regularization to the objective in problem (29). Let

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ & & & \cdots & \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \quad (58)$$

be an $(N_c - 1) \times N_c$ first-order difference matrix. As phase noise does not change significantly between consecutive samples, we can use the regularizer $\|\mathbf{D}\mathbf{u}\|^2$ and solve the following regularized estimation problem:

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} && \mathbf{u}^H \text{Diag}(\mathbf{y})^H \mathbf{V} \text{Diag}(\mathbf{y}) \mathbf{u} + \mu \|\mathbf{D}\mathbf{u}\|^2 \\ & \text{subject to} && |u_n| = 1, n = 0, \dots, N_c - 1, \end{aligned} \quad (59)$$

where μ is a given regularization parameter. To solve the regularized problem with the MM framework, the previous procedure can be followed to obtain the modified MM update, which is given by

$$\begin{aligned} \mathbf{u}^{(t+1)} = \exp & \left[j \arg \left(\text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y}) \mathbf{u}^{(t)} \right) \right. \\ & \left. + \mu \left(\lambda_{\max}(\mathbf{D}^H \mathbf{D}) \mathbf{I}_{N_c} - \mathbf{D}^H \mathbf{D} \right) \mathbf{u}^{(t)} \right]. \end{aligned} \quad (60)$$

Compared with dimensionality reduction method, solving the regularized problem saves the trouble of choosing sampling length N , but still the value of regularization parameter μ needs to be properly chosen. In practice, the obtained estimates are robust to μ within a large range of value.

E. Computational Issues and Analysis of Convergence

In the initialization, both LQM and TQM need to compute \mathbf{B} . Since \mathbf{S} is diagonal, computing $\mathbf{S}^H \mathbf{S}$ needs N_c complex multiplications. Notice that $\tilde{\mathbf{F}}$ consists of the first L ($L \ll N_c$) columns of unitary DFT matrix \mathbf{F} . And using FFT to compute N_c -point DFT needs $(N_c/2) \log_2 N_c$ complex multiplications and $N_c \log_2 N_c$ complex additions. Then $\tilde{\mathbf{F}}^H \mathbf{S}^H \tilde{\mathbf{S}} \tilde{\mathbf{F}}$, which is $L \times L$ Toeplitz, can be calculated with less than $N_c + N_c \log_2 N_c$ complex multiplications and $2N_c \log_2 N_c$ complex additions. Computing the inverse of this $L \times L$ matrix is an easy task (L is the number of the channel taps, usually around the order of 10), which requires $\mathcal{O}(L^2)$ arithmetical

TABLE I
COMPUTATIONAL COMPLEXITY OF PROPOSED ALGORITHMS FOR COMPUTING PHASE NOISE

Algorithms	Arithmetical Operations			
	Initialization		Update	
	×	+	×	+
TQM			N_c^2	$N_c(N_c - 1)$
LQM	$< N_c + N_c \log_2 N_c +$		$N_c^2 + N_c$	N_c^2
TQM + PCT	$2L \left(\frac{L}{2} \log_2 L + N_c - L \right) + 2N_c L +$	$< 2N_c \log_2 N_c +$	N^2	$N(N - 1)$
LQM + PCT	$2N_c^2 + N_c^2 \log_2 N_c + \mathcal{O}(L^2) \ddagger$	$2L^2 \log_2 L + 2N_c^2 \log_2 N_c \clubsuit$	$N^2 + 2N$	$N^2 + N$
TQM + Reg			$N_c^2 + N_c$	$N_c^2 + 3N_c - 2$

\ddagger Usually $L \ll N_c$. And LQM needs additional N_c complex multiplications and N_c real additions to compute $\|\mathbf{y}\|_\infty^2 \mathbf{1} - |\mathbf{y}|^2$.
 \clubsuit When PCT is applied, additional $N_c^2 - N_c N^2$ and $N_c^2 - N_c N^2 + N_c - N$ complex additions are needed for TQM and LQM, respectively.

TABLE II
CPU TIME OF DIFFERENT ALGORITHMS WITH $\Delta f_{3\text{dB}} = 5000$ HZ, SNR = 35 dB, AND 500 MONTE CARLO SIMULATIONS[†]

Algorithms	CPU Time (s)					
	$N_c = 512$			$N_c = 1024$		
	Mean	Min	Max	Mean	Min	Max
PNC + PCT (<i>benchmark</i>)	0.4782	0.4297	0.5952	0.9139	0.8478	1.2091
AltOpt + PCT (<i>benchmark</i>)	0.6427	0.4915	0.9084	3.1677	2.6017	5.3007
AltMM + PCT (<i>modified benchmark</i>)	0.2299	0.1723	0.3232	0.8926	0.6785	1.3401
TQM (<i>proposed</i>)	0.0169	0.0155	0.0263	0.0845	0.0758	0.1047
LQM (<i>proposed</i>)	0.2467	0.0468	0.3667	1.5480 [‡]	0.3011	2.1801
TQM + PCT (<i>proposed</i>)	0.0125	0.0121	0.0145	0.0491	0.0483	0.0542
LQM + PCT (<i>proposed</i>)	0.0159	0.0154	0.0211	0.0637	0.0626	0.0700
TQM + opt-PCT (<i>proposed</i>)	0.0784	0.0753	0.1068	0.3694	0.3596	0.4090
TQM + Reg, $\mu = 50$ (<i>proposed</i>)	0.0198	0.0168	0.0303	0.1296	0.0915	0.2852
TQM + Reg, $\mu = 100$ (<i>proposed</i>)	0.0215	0.0169	0.0386	0.1528	0.0930	0.4229

[†] CPU time measured for each algorithm includes only computing phase noise estimate. For our proposed algorithms, Fast Fourier transform (FFT) is employed to improve computational efficiency.

[‡] LQM costs more CPU time when N_c is large. But some efficient acceleration schemes can be used to boost the convergence rate, e.g., the SQUAREM method [38].

operations [35]. Furthermore, less computation will be required when the pilot symbols have constant modulus, e.g., using M -QAM signals [8], [22], for which $\mathbf{S}^H \mathbf{S} = \alpha \mathbf{I}_L$ for some constant α and \mathbf{B} is thus much simplified. To compute \mathbf{B} , we still need additional FFT for the left and right matrix multiplications, each of which involves $N_c L + ((L/2) \log_2 L + N_c - L) L$ complex multiplications and $L^2 \log_2 L$ complex additions. In the update of TQM, $\text{Diag}(\mathbf{y})^H \mathbf{F}^H \mathbf{B} \mathbf{F} \text{Diag}(\mathbf{y})$ is fixed and thus needs to be computed only once, which requires $2N_c^2 + N_c^2 \log_2 N_c$ complex multiplications and $2N_c^2 \log_2 N_c$ complex additions. The summary of the computational complexity for computing phase noise estimate is given in Table I. The overall complexity is at the order of $\mathcal{O}(N_c^2 \log_2 N_c)$ for the initialization and $\mathcal{O}(N_c)$ for each update. For TQM, the algorithm usually converges in practice within around 10 iterations. Faster convergence speed can also be achieved by applying accelerations; see Table II. Once a phase noise estimate is obtained, channel estimate (26) can be computed with N_c complex multiplications.

Compared with some existing methods, our proposed algorithms requires almost the same or even less computational complexity. For instance, the state-of-the-art PNC in [22] solves an SDP, which requires a complexity of $\mathcal{O}(N^{3.5} \log_2(\frac{1}{\epsilon}))$ using efficient solver SeDuMi [36]. An alternating optimization

method proposed in [19] requires a per-update complexity of $\mathcal{O}(N_c^3)$. And a non-iterative method in [21] requires a complexity of $\mathcal{O}(N^3)$ apart from the computation of \mathbf{B} .

According to (41), the obtained series of points $\{\mathbf{u}^{(t)}\}$ of LQM and TQM is non-increasing. To further show the proposed algorithms converge to a stationary point, we first define a first-order optimality condition for minimizing a smooth function subject to an arbitrary constraint [37].

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

$$\nabla f(\mathbf{x}^*)^T \mathbf{y} \geq 0 \quad \forall \mathbf{y} \in T_{\mathcal{X}}(\mathbf{x}^*), \quad (61)$$

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

With Proposition 1, the convergence of our proposed algorithms is guaranteed, i.e., the limit point of $\{\mathbf{u}^{(t)}\}$ is a stationary point. A similar proof can be found in [33, Th. 5].

V. SIMULATION RESULTS

In simulations, we consider a Rayleigh fading channel of length $L = 10$, where each tap is independently distributed with exponentially decreasing power of rate 0.7 and channel noise is circularly symmetric complex Gaussian with $\sigma = 0.1$.

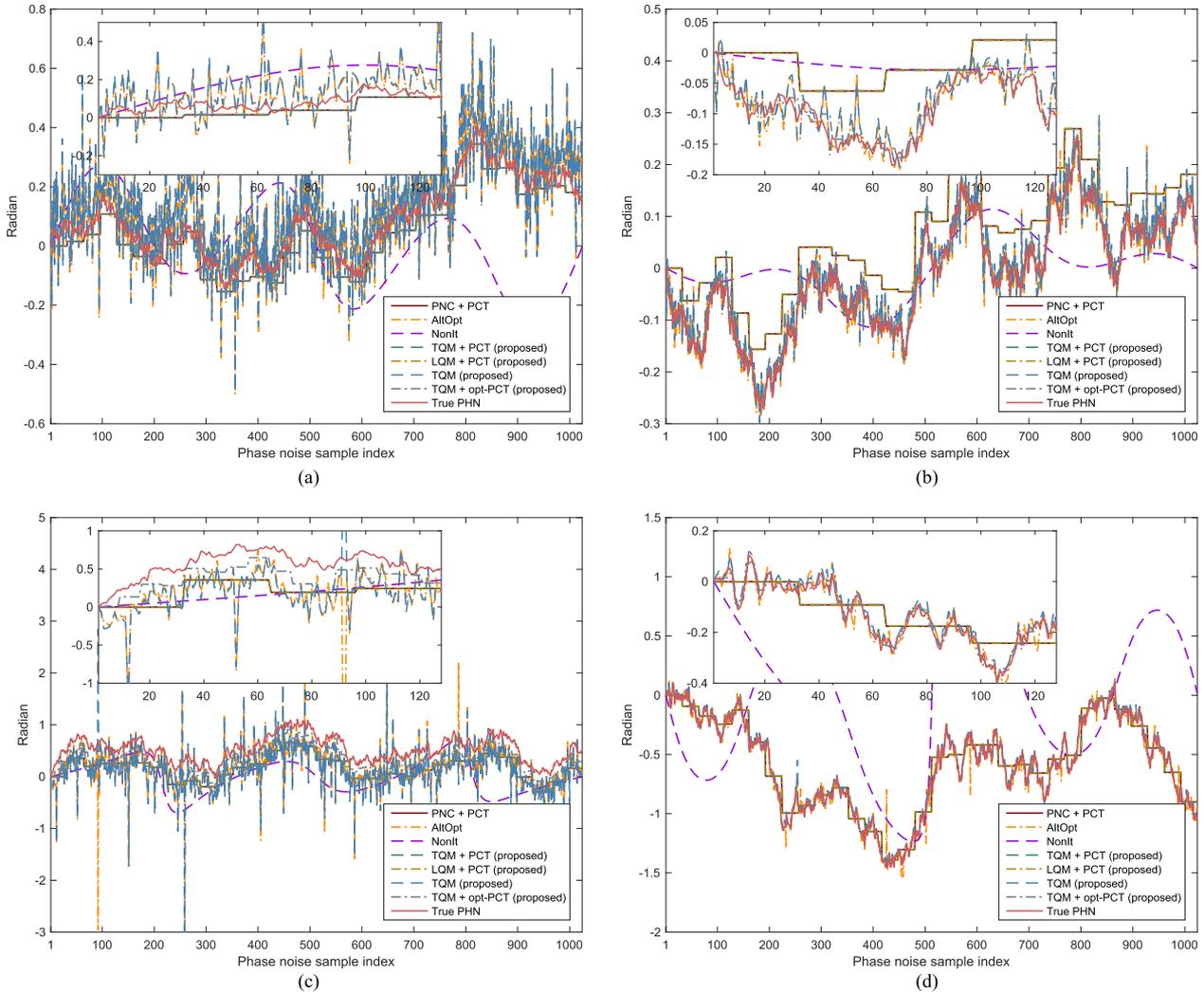


Fig. 1. Four instances of the estimated phase noise θ . $N_c = 1024$. Where PCT is applied, $N = 32$. (a) SNR = 15 dB, $\Delta f_{3\text{dB}} = 500$ Hz, (b) SNR = 35 dB, $\Delta f_{3\text{dB}} = 500$ Hz, (c) SNR = 15 dB, $\Delta f_{3\text{dB}} = 5000$ Hz, (d) SNR = 35 dB, $\Delta f_{3\text{dB}} = 5000$ Hz².

Transmitted OFDM symbols are generated randomly (assumed known to receiver) with distribution $\mathcal{CN}(\mathbf{0}, 2\mathbf{I})$. To apply dimensionality reduction,² we use PCT with the reduced dimension indicated by the value of N (51)–(52). Throughout this section we choose $N = 32$ if a fixed PCT is applied.

The following methods are considered in our simulations. PNC, as the benchmark, is the algorithm proposed in [22], where phase noise is estimated in the frequency domain by solving an SDP (23). A non-iterative method NonIt proposed in [21] is also considered. Method AltOpt refers to the alternating optimization algorithm proposed in [19]. A modified version AltMM, based on the MM, is also compared. Specifically, we have modified AltOpt to take into account the phase noise constraint in each phase noise estimate update as opposed to the original algorithm; see Appendix for details. TQM and LQM are two MM-based algorithms we have proposed to

solve the time-domain problem (29). From a set of specified PCTs with different values of N , opt-PCT is defined as the one that gives the minimal BIC. In particular, opt-PCT is selected from PCTs with $N \in \{2^5, 2^6, \dots, 2^{\log_2 N_c}\}$. TQM + Reg is the regularization-based method as an alternative to dimensionality reduction and to accelerate the convergence speed, we have employed SQUAREM method [38]. For comparison, Ignore PHN and Exact PHN are also included, where phase noise is ignored and the exact phase noise is used, respectively, when estimating channel impulse response. Whenever necessary, an algorithm is initiated with an all-one vector and regarded converged when the squared ℓ_2 -norm of the difference between two consecutive iterates is no larger than 10^{-8} . The maximum number of iterations allowed toward the convergence is 500. All simulations were run in MATLAB on a PC with a 3.20 GHz i5-4570 CPU and 8 GB RAM.

A. Phase Noise and Channel Estimation

In this section, we define phase noise θ as a Wiener process [8], [19], [22]. The baseband sampling rate is $f_s = 20$ MHz;

²The non-iterative method NonIt [21] has been favored for slow-varying phase noise with a small number of OFDM subcarriers. For moderately fast-varying phase noise or when the number of subcarriers is large, the estimates obtained by NonIt are much worse than other methods; also see [17].

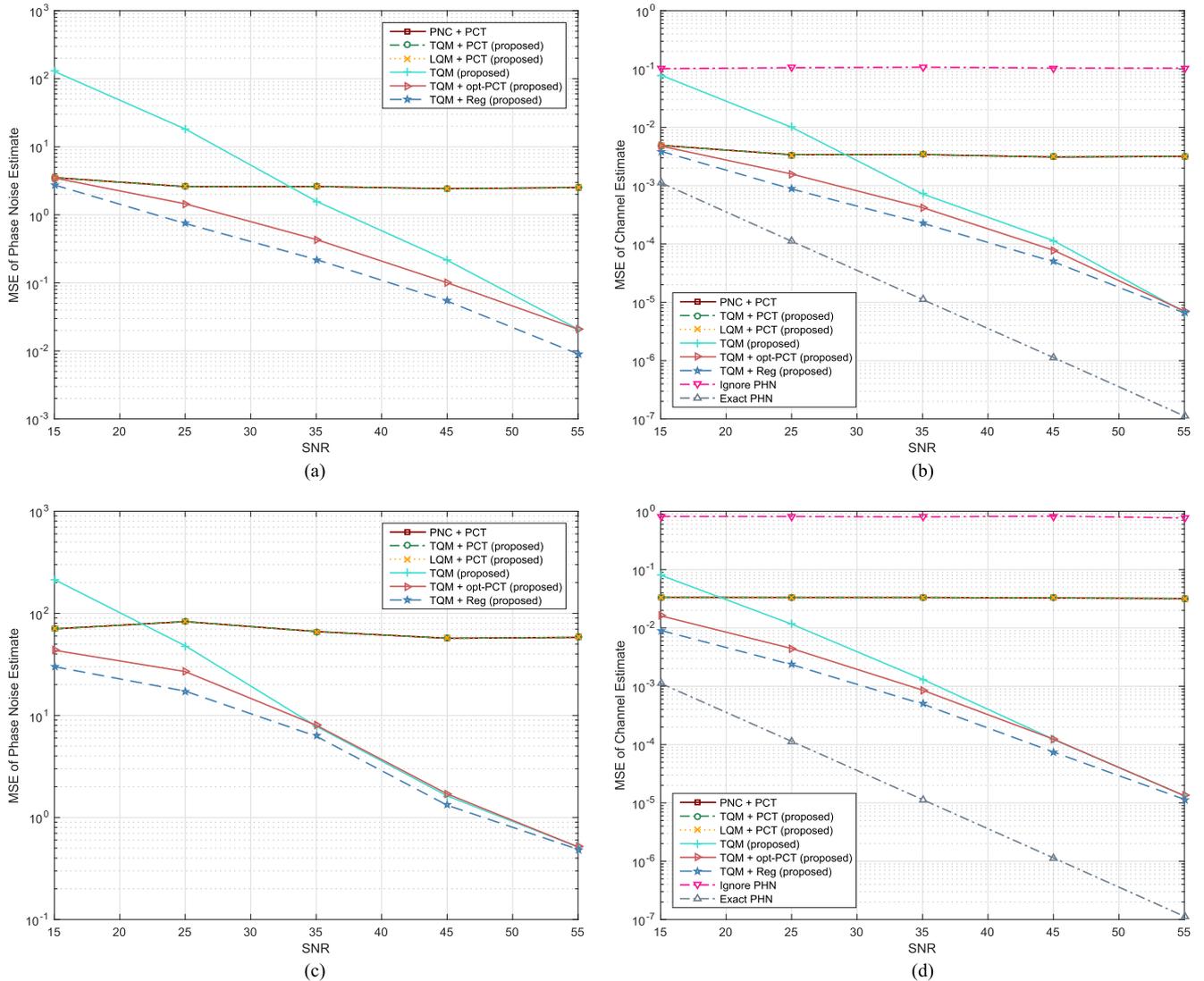


Fig. 2. Joint estimation of phase noise and channel under different values of SNR with $N_c = 1024$ and 2000 Monte Carlo simulations: (a) averaged MSE of phase noise estimate, $\Delta f_{3\text{dB}} = 500$ Hz (b) averaged MSE of channel estimate, $\Delta f_{3\text{dB}} = 500$ Hz, (c) averaged MSE of phase noise estimate, $\Delta f_{3\text{dB}} = 5000$ Hz, (d) averaged MSE of channel estimate, $\Delta f_{3\text{dB}} = 5000$ Hz.

3-dB bandwidth $\Delta f_{3\text{dB}}$ of phase noise is 500 Hz or 5000 Hz; assuming CFO has been fixed, i.e., $\theta_{0=0}$, phase noise is generated with

$$\theta_n - \theta_{n-1} \sim \mathcal{N}\left(0, \sqrt{\frac{2\pi\Delta f_{3\text{dB}}}{f_s}}\right), \quad (62)$$

for $n = 1, \dots, N_c - 1$. We first show four instances of phase noise estimates to provide an intuitive idea of how different algorithms perform, and then compare the resultant phase noise and channel estimation errors by Monte Carlo simulations.

Figs. 1(a)–1(d) show phase noise estimates under four different scenarios. In all cases, PNC, TQM, and LQM yield the same estimate when the a fixed PCT is applied. NonIt, the non-iterative method, solves the original phase noise estimation problem, assuming that the magnitude of phase noise is small, with a first-order Taylor approximation. Also assuming

that phase noise is low pass, it recovers phase noise using only several (3 in our simulations) low-pass spectral components. Obviously, the obtained estimates by NonIt are only rough approximations compared with the true phase noise. Due to this reason, PNC is proposed in [22] to both take spectral phase noise constraints into account and also make use of more spectral components.

1) *Small Phase Noise and Low SNR*: In the small phase noise case with $\Delta f_{3\text{dB}} = 500$ Hz, as Fig. 1(a) shows, using the given PCT results in a staircase-like estimate; loose though it may seem, it is actually beneficial when SNR is limited, owing to the fundamental underdetermined issue of the original problem. In fact, TQM with opt-PCT provides the same estimate as that of the benchmark—opt-PCT in this case is \mathbf{T}_{32} . In contrast, TQM without PCT and AltOpt produce estimates with many undesired peaks associated with larger MSE. It implies, therefore, that with small phase noise and low SNR, dimensionality

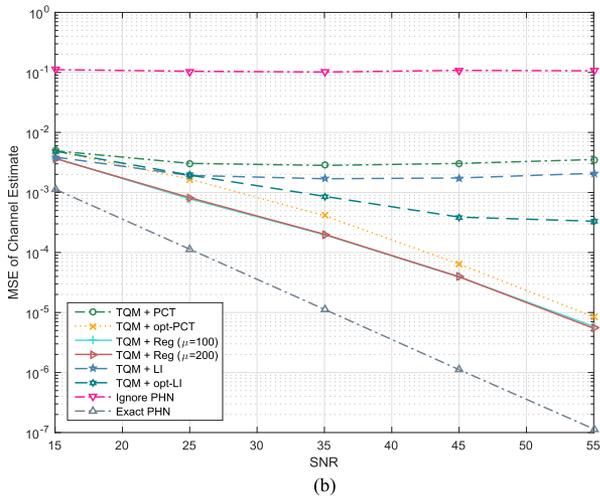
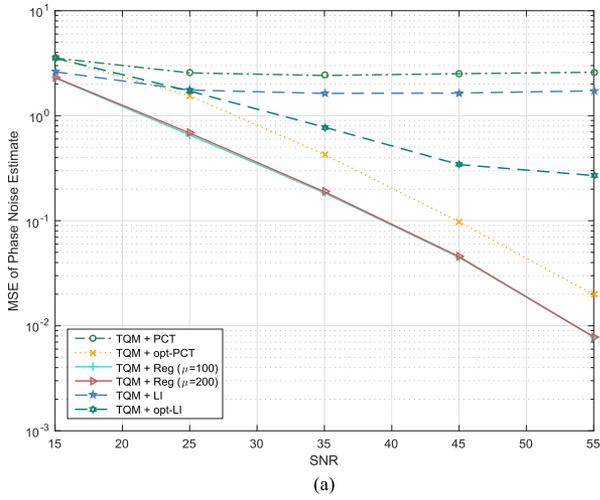


Fig. 3. Joint estimation of phase noise and channel with different values of SNR, $N_c = 1024$, $\Delta f_{3\text{dB}} = 500$ Hz, and 500 Monte Carlo simulations: (a) averaged MSE of phase noise estimate, (b) averaged MSE of channel estimate.

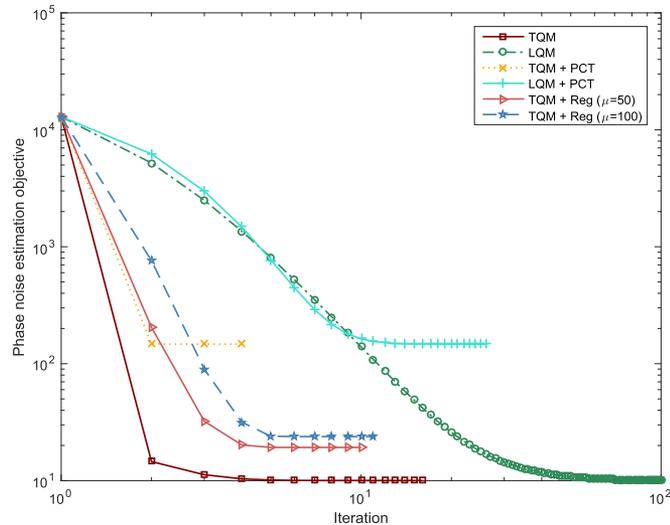


Fig. 4. Convergence of TQM and LQM. $N_c = 1024$. $\Delta f_{3\text{dB}} = 500$ Hz.

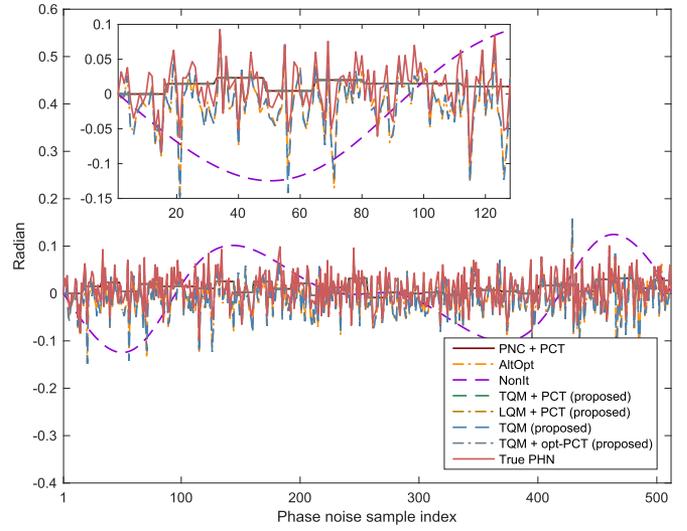


Fig. 5. Gaussian phase noise estimation with $N_c = 512$ and SNR = 35 dB.

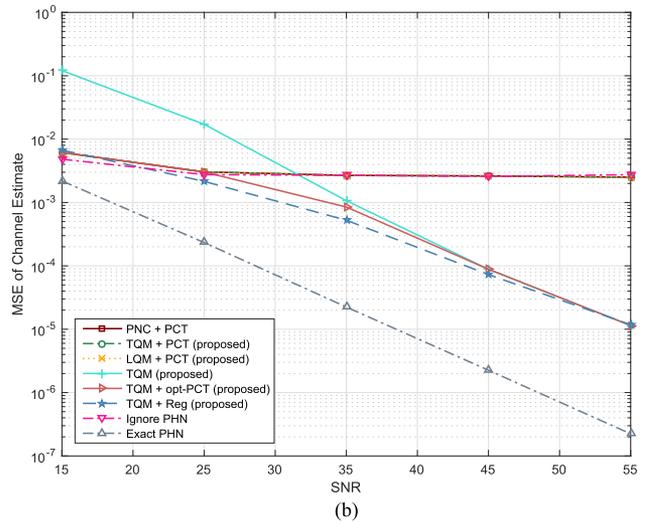
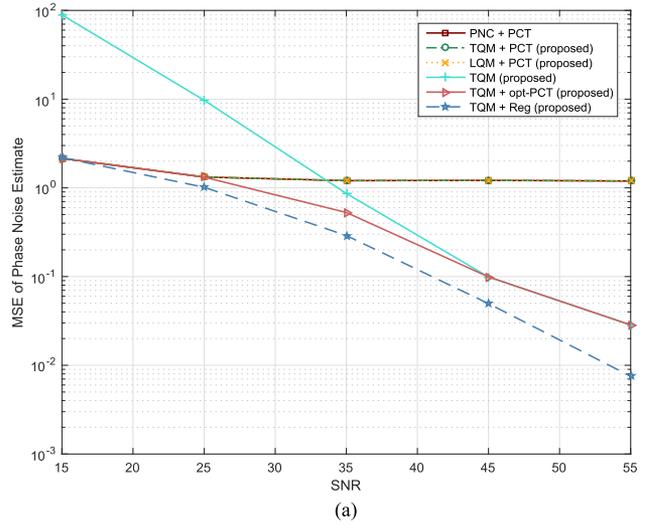


Fig. 6. Joint estimation of Gaussian phase noise and channel with different values of SNR, $N_c = 512$, and 500 Monte Carlo simulations: (a) averaged MSE of phase noise estimate, (b) averaged MSE of channel estimate.

reduction is recommended in order to achieve a relatively better performance.

2) *Small Phase Noise and High SNR*: A particular example for this case is shown in Fig. 1(b). With the given PCT, we can see that PNC, TQM, and LQM are still able to produce the same good estimate by and large, the resulting MSE of which is around 6.98. AltOpt offers a relatively better estimate with MSE 0.6849, which is comparable to TQM without PCT whose MSE is 0.7489. TQM with opt-PCT (opt-PNC is \mathbf{T}_{256} here) outperforms other opponents though, with the corresponding MSE of 0.2362, which implies that dimension should not be reduced too much with high SNR. It should be mentioned that this will pose a challenge to the benchmark PNC because it needs to deal with an SDP with size of 512 or even larger. More details will be illustrated in Section V-B.

3) *Large Phase Noise and Low SNR*: A major difference in this scenario from the previous one is that TQM without PCT yields a phase noise estimate with many undesired peaks and so does AltOpt; see Fig. 1(c). The reason is that low SNR renders the original problem more susceptible to the underdetermined issue—the same as the case of small phase noise and low SNR in Fig. 1(a). Nonetheless, TQM with opt-PCT gives a very good estimate outperforming other opponents (opt-PCT in this case is \mathbf{T}_{128}).

4) *Large Phase Noise and High SNR*: In this last example, except NonIt, all the other algorithms yield nearly good estimates as Fig. 1(d) shows. Fixed PCT for PNC, TQM, and LQM, however, still provide a relatively loose result that could have been improved with available high SNR, the MSE of which are 9.9862, 9.9862, and 9.9870, respectively. AltOpt yields an estimate with smaller MSE 1.8928. We can also expect that TQM without PCT and TQM with opt-PCT perform better, the resulting MSE of which are 0.6938 and 0.5352, respectively. Opt-PCT in this example is \mathbf{T}_{512} .

Since phase noise estimate by NonIt is much worse than other methods and AltOpt and AltMM have MSE of phase noise no better than TQM, we will only compare our proposed algorithms with PNC with respect to MSE. For comparison of computation time for each algorithm, see Table II.

Figs. 2(a)–2(d) show averaged MSE of phase noise and channel estimates with 500 Monte Carlo trials. From Fig. 2(a) and 2(c), we see that PNC, TQM, and LQM are comparable to each other with dimensionality reduction when SNR is low. In contrast, TQM without PCT can provide much better phase noise estimates for high enough SNR. A similar result can be found for channel estimation in Fig. 2(b) and 2(d). But TQM produces better estimates with opt-PCT than without PCT. Even though the benchmark PNC can deal with larger SNR, the computational burden is prohibitive, not to say the additional computational issues; see the following remark and an illustration of CPU time consumed by each algorithm in Section V-B. Using regularization, TQM can further reduce MSE compared with TQM with PCT. Here, we choose regularization parameter $\mu = 50$ for $\Delta f_{3\text{dB}} = 500$ Hz and $\mu = 10$ for $\Delta f_{3\text{dB}} = 5000$ Hz. But for a given $\Delta f_{3\text{dB}}$, the performance of resulting estimates is robust to the choice of μ . See Fig. 3(a)–3(b) for more details. The LI is the interpolation matrix proposed in [19] and opt-LI, similar to

opt-PCT, is obtained with BIC. For LI to be properly defined, the sampling length N is chosen from $\{34, 94, 342, 1024\}$, which is more restricted than PCT. For fixed LI, we choose $N = 34$.

Remark 2: PNC is proposed in [22], where the original problem is reformulated as an SDP by using the S-procedure. The authors only prove the equivalence (strong duality) between the reduced problem and its SDP reformulation. For the original problem, however, the strong duality has not been established. From simulations, the resultant estimate of frequency-domain phase noise vector is not a reasonably good solution that satisfies the spectral constraint. Also, solving an SDP only gives an intermediate solution that requires an additional eigendecomposition step. PNC can thus easily fall within infeasibility and singularity issues when dimension is not reduced enough. These issues, however, have not been addressed in [22].

B. Algorithm Convergence

In this section, we first present an example of convergence properties of our proposed algorithms, and then give a comparison of CPU time consumed by each algorithm. Convergence criteria defined previously apply here as well.

Fig. 4 demonstrates convergence of four methods. TQM and LQM converge to the same optimal solution with the same initialization. TQM, however, converges remarkably faster than LQM, within twenty iterations or fewer. This is because TQM employs a much tighter majorization function than LQM. Thus we adopt TQM with opt-PCT in previous simulations to achieve the same performance with respect to estimation error and to save computation time at the same time. On the other hand, as shown in previous examples, applying PCT in the case of large phase noise and high SNR causes loss of quality in the obtained estimates, which is substantiated by the fact that without PCT, much lower objective value can be achieved. Using different values of regularization parameters, TQM converges to different objective values with comparable computation time. The resulting estimation errors, however, are almost the same as Figs. 2(a)–2(b) have shown.

A further comparison in terms of the computation time among our proposed algorithms and the benchmarks is provided in Table II. In the case with PCT applied, our proposed algorithms outperform PNC, AltOpt, and AltMM by saving much time and at the same time achieve the same MSE as shown in the previous examples. TQM gives as the same estimate as LQM but is much more efficient owing to the tighter majorization function employed. TQM with regularization is also computationally efficient with an even further improvement in the resulting estimation errors.

C. Gaussian Phase Noise Estimation

Phase noise generated in a phase-locked loop is modeled as a Gaussian process [18]. In this example, the number of subcarriers is $N_c = 512$ with baseband sampling rate $f_s = 20$ MHz. The standard deviation θ_{rms} of phase noise generated by a phase-locked voltage controlled oscillator is 2 degrees. The single-pole butterworth filter with 3-dB bandwidth $\Delta f_{3\text{dB}} = 100$ Hz is adopted so that the covariance matrix of

phase noise is

$$\mathbf{C}_{i,j} = \left(\frac{\pi \theta_{\text{rms}}}{180} \right)^2 e^{\frac{-2\pi \Delta f_{3dB} |i-j|}{f_s}}. \quad (63)$$

An instance of Gaussian phase noise and its estimates is shown in Fig. 5. As its name indicates, Gaussian phase noise will not drift away too much like Wiener phase noise. Similar to Fig. 1(b) and 1(d), large dimensionality reduction induces considerable loss in the obtained estimates; with PCT, MSE for PNC, TQM, and LQM are 1.4049, 1.4049, and 1.4024, respectively. TQM without PCT and TQM with opt-PCT achieve the best performance in this example, both of which have the same MSE of 0.1556. And opt-PCT is just an identity matrix, which causes no dimension to be reduced. Figs. 6(a)–6(b) show MSE of the obtained estimates of Gaussian phase noise and channel, respectively. Still, our proposed algorithms can provide much better estimates as in the previous examples with Wiener phase noise.

VI. CONCLUSION

We have proposed efficient algorithms for the joint phase noise and channel estimation in OFDM. The algorithms are devised based on the majorization-minimization technique and apply to two canonical models of phase noise—Wiener process and Gaussian process. To properly address the underdetermined nature in the original estimation problem, dimensionality reduction and regularization have been proposed with similar MM algorithms provided. The simulation results have shown that when the same dimensionality reduction is employed, our proposed algorithms achieve the same MSE as that of the benchmark but consume much less time. By further selecting the optimal dimensionality reduction with BIC or imposing appropriate regularization, our proposed algorithms produce significantly better estimates for moderate SNR without demanding much additional computation time. It is expected that in modern applications of OFDM, where a large number of subcarriers are deployed, the advantage of our methods should be outstanding.

APPENDIX

ALTERNATING OPTIMIZATION WITH THE MM

In the following, the constraint of phase noise is taken into account and the alternating optimization scheme is correspondingly modified.

With the unimodular constraint for $\mathbf{c} = e^{j\theta}$, we have $(\Phi_\phi^{(i-1)})^H \Phi_\phi^{(i-1)} = N_c \mathbf{I}$ (14). And the channel estimate is updated by

$$\hat{\mathbf{h}}^{(i)} = (N_c \check{\mathbf{F}}^H \mathbf{S}^H \mathbf{S} \check{\mathbf{F}})^{-1} \check{\mathbf{F}}^H \mathbf{S}^H (\Phi_\phi^{(i-1)})^H \mathbf{r}. \quad (64)$$

Substitute (64) into the objective of problem (15), and the following problem is obtained:

$$\underset{\mathbf{c}: |c_n|=1, n=1, \dots, N_c}{\text{minimize}} \quad \|\mathbf{r} - \mathbf{P}\mathbf{F}\mathbf{c}\|^2, \quad (65)$$

where $\mathbf{P} = \text{circ}(\check{\mathbf{S}}\hat{\mathbf{F}}\hat{\mathbf{h}}^{(i)})$. Instead of updating \mathbf{c} by the least-squares solution (17), the MM method can be used to solve

Algorithm 3: Phase Noise Estimation by Alternating Minimization and the MM.

- 1: Set $i = 1$ and initialize $\hat{\mathbf{c}}^{(0)} = e^{j\theta_0}$.
 - 2: **repeat**
 - 3: $\Phi_\phi^{(i-1)} = \text{circ}(\mathbf{F}\hat{\mathbf{c}}^{(i-1)})$
 - 4: $\hat{\mathbf{h}}^{(i)} = (N_c \check{\mathbf{F}}^H \mathbf{S}^H \mathbf{S} \check{\mathbf{F}})^{-1} \check{\mathbf{F}}^H \mathbf{S}^H (\Phi_\phi^{(i-1)})^H \mathbf{r}$
 - 5: $\mathbf{P} = \text{circ}(\check{\mathbf{S}}\hat{\mathbf{F}}\hat{\mathbf{h}}^{(i)})$
 - 6: $\lambda = N_c \|\mathbf{F}^H (\check{\mathbf{S}}\hat{\mathbf{F}}\hat{\mathbf{h}}^{(i)})\|_\infty^2$
 - 7: $t = 0$, and initialize $\mathbf{c}^{(0)}$
 - 8: **repeat**
 - 9: $\mathbf{a}^{(t+1)} = \mathbf{F}^H \mathbf{P}^H \mathbf{r} + \lambda \mathbf{c}^{(t)} - \mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F} \mathbf{c}^{(t)}$
 - 10: $\mathbf{c}^{(t+1)} = e^{j \arg \mathbf{a}^{(t+1)}}$
 - 11: **until** convergence
 - 12: $\hat{\mathbf{c}}^{(i)} = \mathbf{c}^{(t+1)}$
 - 13: $t \leftarrow t + 1$
 - 14: **until** convergence.
-

problem (65). The majorization can be obtained as follows [33, Lemma 1]:

$$\|\mathbf{r} - \mathbf{P}\mathbf{F}\mathbf{c}\|^2 = \mathbf{r}^H \mathbf{r} - 2\text{Re}\{\mathbf{r}^H \mathbf{P}\mathbf{F}\mathbf{c}\} + \mathbf{c}^H \mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F} \mathbf{c} \quad (66)$$

$$\begin{aligned} &\leq \mathbf{r}^H \mathbf{r} - 2\text{Re}\{\mathbf{r}^H \mathbf{P}\mathbf{F}\mathbf{c}\} + \lambda \mathbf{c}^H \mathbf{c} \\ &\quad + 2\text{Re}\left\{(\mathbf{c}^{(t)})^H (\mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F} - \lambda \mathbf{I}) \mathbf{c}\right\} \\ &\quad + (\mathbf{c}^{(t)})^H (\lambda \mathbf{I} - \mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F}) \mathbf{c}^{(t)}. \end{aligned} \quad (67)$$

To obtain a good majorization function, we can choose λ as

$$\lambda = \lambda_{\max}(\mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F}) \quad (68)$$

$$= N_c \left\| \mathbf{F}^H (\check{\mathbf{S}}\hat{\mathbf{F}}\hat{\mathbf{h}}^{(i)}) \right\|_\infty^2, \quad (69)$$

where Lemma 1 is applied to compute the maximum eigenvalue of \mathbf{P} . Minimizing (67) results in the update of \mathbf{c} :

$$\mathbf{c}^{(t+1)} = e^{j \arg \mathbf{a}^{(t+1)}}, \quad (70)$$

where $\mathbf{a}^{(t+1)} = \mathbf{F}^H \mathbf{P}^H \mathbf{r} + \lambda \mathbf{c}^{(t)} - \mathbf{F}^H \mathbf{P}^H \mathbf{P} \mathbf{F} \mathbf{c}^{(t)}$. The complete procedure is described in Algorithm 3. Algorithm 3 can also be readily modified to incorporate PCT, for which a similar majorization approach can be followed.

REFERENCES

- [1] T. Pollet, M. V. Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 191–193, Feb. 1995.
- [2] L. Tomba, "On the effect of Wiener phase noise in OFDM systems," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 580–583, May 1998.
- [3] A. G. Armada, "Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM)," *IEEE Trans. Broadcast.*, vol. 47, no. 2, pp. 153–159, Jun. 2001.
- [4] C. Muschallik, "Influence of RF oscillators on an OFDM signal," *IEEE Trans. Consum. Electron.*, vol. 41, no. 3, pp. 592–603, Aug. 1995.
- [5] T. H. Lee and A. Hajimiri, "Oscillator phase noise: A tutorial," *IEEE J. Solid-State Circuits*, vol. 35, no. 3, pp. 326–336, Mar. 2000.

- [6] L. Piazzi and P. Mandarini, "Analysis of phase noise effects in OFDM modems," *IEEE Trans. Commun.*, vol. 50, no. 10, pp. 1696–1705, Oct. 2002.
- [7] K. Nikitopoulos and A. Polydoros, "Phase-impairment effects and compensation algorithms for OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 698–707, Apr. 2005.
- [8] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1607–1616, Aug. 2007.
- [9] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Performance analysis of OFDM with Wiener phase noise and frequency selective fading channel," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1321–1331, May 2011.
- [10] P. Mathecken, T. Riihonen, N. N. Tchamov, S. Werner, M. Valkama, and R. Wichman, "Characterization of OFDM radio link under PLL-based oscillator phase noise and multipath fading channel," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1479–1485, Jun. 2012.
- [11] P. Robertson and S. Kaiser, "Analysis of the effects of phase-noise in orthogonal frequency division multiplex (OFDM) systems," in *Proc. IEEE Int. Conf. Commun.*, Seattle, WA, USA, Jun. 1995, vol. 3, pp. 1652–1657.
- [12] K. Sathananthan and C. Tellambura, "Performance analysis of an OFDM system with carrier frequency offset and phase noise," in *Proc. IEEE 54th Veh. Technol. Conf.*, 2001, vol. 4, pp. 2329–2332.
- [13] S. Wu and Y. Bar-Ness, "OFDM channel estimation in the presence of frequency offset and phase noise," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AK, USA, May 2003, vol. 5, pp. 3366–3370.
- [14] S. Wu and Y. Bar-Ness, "OFDM systems in the presence of phase noise: Consequences and solutions," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1988–1996, Nov. 2004.
- [15] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, no. 10, pp. 2908–2914, Oct. 1994.
- [16] V. Syrjala, M. Valkama, N. N. Tchamov, and J. Rinne, "Phase noise modelling and mitigation techniques in OFDM communications systems," in *Proc. Wireless Telecommun. Symp.*, Apr. 2009, pp. 1–7.
- [17] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Constrained phase noise estimation in OFDM using scattered pilots without decision feedback," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2348–2362, May 2017.
- [18] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint estimation of channel response, frequency offset, and phase noise in OFDM," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3542–3554, Sep. 2006.
- [19] Q. Zou, A. Tarighat, and A. H. Sayed, "Compensation of phase noise in OFDM wireless systems," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5407–5424, Nov. 2007.
- [20] R. Carvajal, J. C. Aguero, B. I. Godoy, and G. C. Goodwin, "EM-based maximum-likelihood channel estimation in multicarrier systems with phase distortion," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 152–160, Jan. 2013.
- [21] P. Rabiei, W. Namgoong, and N. Al-Dhahir, "A non-iterative technique for phase noise ICI mitigation in packet-based OFDM systems," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5945–5950, Nov. 2010.
- [22] P. Mathecken, T. Riihonen, S. Werner, and R. Wichman, "Phase noise estimation in OFDM: Utilizing its associated spectral geometry," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1999–2012, Apr. 2016.
- [23] *Frame Structure Channel Coding and Modulation for a Second Generation Digital Terrestrial Television Broadcasting System (DVB-T2)*, ETSI Standard EN 302 755 V1.3.1, 2009.
- [24] R. A. Casas, S. L. Biracree, and A. E. Youtz, "Time domain phase noise correction for OFDM signals," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 230–236, Sep. 2002.
- [25] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "OFDM channel estimation in the presence of phase noise and frequency offset by particle filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, USA, Apr. 2007, vol. 3, pp. 289–292.
- [26] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "Monte Carlo methods for channel, phase noise, and frequency offset estimation with unknown noise variances in OFDM systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3613–3626, Aug. 2008.
- [27] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [28] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [29] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, 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] Y. Sun, P. Babu, and D. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2016.
- [32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [33] J. Song, P. Babu, and D. P. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, Apr. 2015.
- [34] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
- [35] W. F. Trench, "An algorithm for the inversion of finite Toeplitz matrices," *J. Soc. Ind. Appl. Math.*, vol. 12, no. 3, pp. 515–522, 1964.
- [36] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11/12, pp. 625–653, 1999.
- [37] D. P. Bertsekas *et al.*, *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Sci., 2003.
- [38] 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.



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 the Hong Kong University of Science and Technology, Hong Kong. His research interests include applications of convex optimization to signal processing, machine learning, and financial engineering.

Prabhu Babu received the Ph.D. degree in electrical engineering from Uppsala University, Uppsala, Sweden, in 2012. From 2013 to 2016, he was a Postdoctoral Fellow with the Hong Kong University of Science and Technology. He is currently with the Center for Applied Research in Electronics, Indian Institute of Technology Delhi, New 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 currently 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. Since 2013, he has been a Fellow of the Institute for Advance Study, HKUST. He had previously held several research appointments, namely, at King's College London, London, UK; Stanford University, Stanford, CA, USA; Telecommunications Technological Center of Catalonia, Barcelona, Spain; Royal Institute of Technology (KTH), 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.

Dr. Palomar received a 2004/06 Fulbright Research Fellowship, the 2004 and 2015 (coauthor) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015–2016 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.