Widely Linear Estimation of Oscillator Phase Noise with Rank Reduction

Aamir Ishaque and Gerd Ascheid
Institute for Communication Technologies and Embedded Systems, RWTH Aachen University, Germany
Email: {ishaque,ascheid}@ice.rwth-aachen.de

***Abstract*—In this paper, we present a novel algorithm for oscillator phase noise estimation using two key properties of the phase noise process in digital baseband: reduced-dimensional characteristics when expanded with Karhunen-Loeve basis functions and statistical improperness of the phase noise coefficients paving the way for the optimality of widely-linear (WL) estimators. The proposed methods are designed to take full-advantage of the second-order statistics, rank-deficient models and has an attractive trade-off between performance and complexity. To compute rank-reduced WL Wiener filter for highly compressed estimation, two methods are derived and analyzed using either data or parameter subspace reduction. Numerical experiments are presented for practically relevant phase noise distributions and they demonstrate the applicability of the proposed methods. They show that WL estimator can achieve up to 5 dB improvement over the best strictly linear estimator, with the maximum achieved in complexity-favorable low-rank conditions.**

Keywords: Phase noise, Improper signals, Wiener filter, Karhunen-Loeve transform.

I. Introduction and Background

Modern wireless communication systems, especially the multicarrier waveforms, are sensitive to phase noise distortion [1]. It appears as a random phase deviation of the frequency synthesizer over time with respect to (w.r.t) a pure sinusoidal tone. With the emergence of fully integrated solutions for mobile terminals, the issue of phase noise becomes even more pronounced, owing to the inferior fabrication accuracy in cost effective designs. Striving for higher spectral efficiency and throughput, efficient estimation of "mitigable" distortion becomes a key parameter for the deployment of higher-order constellations.

A number of approaches for phase tracking have been proposed, including Fourier transform based approaches [2]–[4], state-space algorithm based on Kalman filter [5] and linear-minimum-mean-square-error (LMMSE) based methods [1], [6]–[8]. The problem of phase noise estimation is more involved, than for example channel estimation, due to its continuous evolution nature. A reduced-rank estimation is indeed superior in many respects. In many applications, high computational complexity is undesirable or model reduction is necessary to enhance robustness against noise at the expense of model bias. Fortunately, a realistic phase noise has been known to be a slowly varying process meaning that a high correlation is expected and a low-rank approximation captures the dominant structure of the linear model [3]. It is worthwhile to point out that low-rank DFT approximations suffer from well-known edge effects [4]. Distinction must be made between dimensionality of the data signal and unknown parameters. In this context, principal component analysis has been widely employed for transformation of the received vectors in blind techniques [9], whereas [10] derived a reduced-rank Wiener filter through parameter compression, calling it the generalized Karhunen-Loeve transform (GKLT).

The real phase noise process $\phi(n)$ corrupts the information-bearing signal through a multiplicative term $\theta(n) = e^{\jmath\phi(n)}$. Unlike [4], [7], [8] and by directly tracking $\theta(n)$, our approach does not put any restriction on the magnitude of phase process $\phi(n)$. This paper studies an enhanced phase noise estimator by exploiting the improperness of $\theta(n)$. A zero-mean complex signal $s(n)$ is said to be improper if its pseudo-covariance $\mathcal{E}\left\{s^2(n)\right\}$ is non-zero. This condition is fulfilled when the real and imaginary parts of a complex process are correlated or have unequal variance. It has been shown that the performance of the linear estimators can be improved if not only the observation $y(n)$ but also the complex conjugate $y^*(n)$ of the observation is used for estimation of the desired signal [11][1]. The resulting schemes are commonly known as "widely linear" (WL) solutions. The useful signal's improperness is beneficial for improving the estimation performance, if there exists statistical dependence between $y(n)$ and $y^*(n)$. In the maximum-likelihood context, the increase in degrees of freedom due to improper signals has been exploited for interference suppression [13]. To the best of our knowledge, the application of WL solution to the phase noise estimation problem is still unexplored.

The rest of this paper is organized as follows. In section II, we describe the reference system and introduce WL processing for phase noise estimation. An algorithm to efficiently implement full-rank WL filter is discussed in section III. In section IV we derive a reduced-rank version for the WL-MMSE filtering. Simulation results are presented in V and conclusions are drawn in the last section.

Notation: $(\cdot)^\dagger$, $\|\cdot\|_F$, $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote pseudo-inverse, Frobenius norm, conjugation, transpose and conjugate transpose operations respectively; $\mathbf{I}$ and $\mathbf{0}$ represent the identity matrix and the null matrices respectively; $\mathrm{Tr}\{\cdot\}$ and $\mathcal{E}\{\cdot\}$ refer to trace and expectation operators respectively; $\jmath = \sqrt{-1}$ and $\odot$ denotes element-wise vector product; subscripts $(\cdot)_r$ and $(\cdot)_i$ represent the real and imaginary parts respectively; $\mathrm{diag}\,[\cdot]$ creates a diagonal matrix from the argument vector.

II. System Model And WL Estimator

Our system assumes a bursty transmission of N symbols and the estimator operates in either data-aided or decision directed fashion. The baseband equivalent of a noisy channel is described in the following relation:

$$\boldsymbol{y}' = \boldsymbol{x} \odot \boldsymbol{\theta}' + \boldsymbol{w} \tag{1}$$

where $\boldsymbol{y}' \triangleq [y'(1), y'(2), \ldots, y'(N)]^T$ and $\boldsymbol{w} \triangleq [w(1), w(2), \ldots, w(N)]^T$ symbolize the channel output and the additive proper white Gaussian noise with distribution $w(n) \sim \mathcal{CN}(0, \sigma_w^2)$ respectively. In the case of fading channels, the system input $\boldsymbol{x} \triangleq [x(1), x(2), \ldots, x(N)]^T$ can be treated as a transmit signal filtered by a channel

[1]The performance equivalence between augmented formulation $\bar{\boldsymbol{x}}(n) = [x(n)\ x^*(n)]^T$ and the alternative approach with processing of the real and imaginary inputs: $\tilde{\boldsymbol{x}}(n) = [x_r(n)\ x_i(n)]^T$ has been well established [12]. We will use ascent $\bar{(\cdot)}$ for the former and $\tilde{(\cdot)}$ for the latter case variable formats.

realization that can either be a priori known or estimated jointly with the considered phase tracking algorithm [5]. Moreover $\boldsymbol{\theta}' \triangleq [e^{\jmath\phi(1)}, e^{\jmath\phi(2)}, \ldots, e^{\jmath\phi(N)}]^T$ denotes the unknown and to-be-estimated phase noise process with functional dependence on zero-mean correlated Gaussian-distributed samples: $\phi(n) \sim \mathcal{N}\left(0, \sigma^2_{\phi_n}\right)$, with the cross-correlation $R_{\phi\phi}(n,m) = \mathcal{E}\left\{\phi(n)\phi(m)\right\}$. The a-priori known covariance functions of $\boldsymbol{\theta}$ are: $C_{\theta\theta}(n,m) = \mathcal{E}\left\{\theta'(n)\theta'^*(m)\right\} - \mu_\theta(n)\mu^*_\theta(m)$ and $C_{\overline{\theta\theta}}(n,m) = \mathcal{E}\left\{\theta'(n)\theta'(m)\right\} - \mu_\theta(n)\mu_\theta(m)$, where $\mathcal{E}\left\{\theta'(n)\theta'^*(m)\right\} = e^{-\left(\sigma^2_{\phi_n}+\sigma^2_{\phi_m}-2R_{\phi\phi}(n,m)\right)/2}$ and $\mu_\theta(n) = \mathcal{E}\left\{e^{\jmath\phi(n)}\right\} = \mathcal{E}\left\{e^{-\jmath\phi(n)}\right\} = e^{-\sigma^2_{\phi_n}/2}$.

In the sequel, we subtract the constant part from (1) i.e., $\boldsymbol{y} = \boldsymbol{y}' - \bar{\boldsymbol{y}} = \boldsymbol{x} \odot \boldsymbol{\theta} + \boldsymbol{w} = \boldsymbol{s} + \boldsymbol{w}$, so that both $\boldsymbol{y}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}' - \mathcal{E}\left\{\boldsymbol{\theta}'\right\}$ are zero-mean random variables. We are interested in estimating $\boldsymbol{\theta}'$ such that its complex-valued estimator has the widely-linear structure:

$$\hat{\boldsymbol{\theta}}' = \underbrace{[\mathbf{F}_1 \quad \mathbf{F}_2]}_{\bar{\mathbf{F}}} \underbrace{\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{bmatrix}}_{\bar{\boldsymbol{y}}} + \mathcal{E}\left\{\boldsymbol{\theta}'\right\} = \hat{\boldsymbol{\theta}} + \boldsymbol{\mu}_\theta \tag{2}$$

or equivalently for $\boldsymbol{\theta}$ estimation,

$$\hat{\boldsymbol{\theta}} = \underbrace{[\mathbf{1} \quad \jmath\mathbf{1}]}_{\mathbf{G}} \underbrace{\begin{bmatrix} \hat{\boldsymbol{\theta}}_r \\ \hat{\boldsymbol{\theta}}_i \end{bmatrix}}_{\hat{\tilde{\boldsymbol{\theta}}}} = \mathbf{G} \underbrace{\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}}_{\tilde{\mathbf{H}}} \underbrace{\begin{bmatrix} \boldsymbol{y}_r \\ \boldsymbol{y}_i \end{bmatrix}}_{\tilde{\boldsymbol{y}}} \tag{3}$$

where $\mathbf{H}_{11} = \mathbf{F}_{1,r} + \mathbf{F}_{2,r}$, $\mathbf{H}_{22} = \mathbf{F}_{1,r} - \mathbf{F}_{2,r}$, $\mathbf{H}_{12} = \mathbf{F}_{2,i} - \mathbf{F}_{1,i}$, and $\mathbf{H}_{21} = \mathbf{F}_{1,i} + \mathbf{F}_{2,i}$. Note that for strictly linear (SL) case, $\mathbf{H}_{11} = \mathbf{H}_{22}$ and $\mathbf{H}_{12} = -\mathbf{H}_{21}$.

It is of particular interest to compare WL solutions against their SL variants. Assuming proper signals $s(n)$ and $w(n)$, and a small phase noise magnitude i.e., $\theta(n) \approx 1 + \jmath\phi(n)$ and $s(n) = \jmath x(n)\phi(n)$, the decisive factors are the two correlation measures: $\mu_r = \mathcal{E}_{\phi,x}\left\{s(n)s_r(n)\right\} = \sigma^2_{\phi_n}\mathcal{E}_x\left\{x_i^2(n)\right\}$ and $\mu_i = \mathcal{E}_{\phi,x}\left\{s(n)s_i(n)\right\} = \sigma^2_{\phi_n}\mathcal{E}_x\left\{x_r^2(n)\right\}$, using the fact that $\mathcal{E}_x\left\{x_r(n)x_i(n)\right\} = 0$. For the special case of M-PSK modulation: $x(n) = \exp(\jmath\varphi(n))$ with $\varphi(n) = 2\pi\frac{m_n}{M} + \varphi_0$, de-rotating $y(n)$ we have $\mu_r = 0$, therefore WL processing simply means that imaginary coordinates: $y_i(n) = \phi(n) + w_i(n)$ construct sufficient statistics. Whereas, for a more general case, $x(n) \in \mathbb{C}$, it is the disparity between μ_r and μ_i that allows WL processing to exhibit performance advantage over SL, given the power limit: $\mu = \mu_r + \mu_i$.

Quantitatively, the WL's performance advantage δ_e^2 can be determined using the approximate diagonalization of the covariance matrices. For a unitary matrix $\mathbf{U}$, if $\mathbf{P}_{yy} = \mathcal{E}\left\{\boldsymbol{y}\boldsymbol{y}^T\right\} = \mathbf{U}\boldsymbol{\Lambda}_\mathbf{p}\mathbf{U}^T$, then $\mathbf{C}_{yy} = \mathcal{E}\left\{\boldsymbol{y}\boldsymbol{y}^H\right\} \approx \mathbf{U}\boldsymbol{\Lambda}_\mathbf{c}\mathbf{U}^H$, giving [14]:

$$\delta_e^2 = \boldsymbol{\rho}^H\left(\mathbf{C}^*_{yy} - \mathbf{P}^*_{yy}\mathbf{C}^{-1}_{yy}\mathbf{P}_{yy}\right)^{-1}\boldsymbol{\rho} \tag{4}$$

$$\approx \boldsymbol{\rho}^H\mathbf{U}^*\boldsymbol{\Lambda}\mathbf{U}^T\boldsymbol{\rho} \tag{5}$$

where the diagonal matrices are defined as $\boldsymbol{\Lambda} = \mathrm{diag}\left[\delta_1, \delta_2, \ldots, \delta_N\right]$, $\boldsymbol{\Lambda}_\mathbf{c} = \mathrm{diag}\left[\delta_{c,1}, \delta_{c,2}, \ldots, \delta_{c,N}\right]$ and $\boldsymbol{\Lambda}_\mathbf{p} = \mathrm{diag}\left[\delta_{p,1}, \delta_{p,2}, \ldots, \delta_{p,N}\right]$. The positive semi-definiteness of the inverse matrix in (4) ensures that $\delta_e^2 \geq 0$. From (5), we can easily show that:

$$\delta_n = \frac{\delta_{c,n}}{\delta^2_{c,n} - \delta^2_{p,n}}. \tag{6}$$

As expected, an increase in the degree of improperness, i.e., higher $\delta_{p,n}$, promises higher MSE gain.

In the following sections, we derive FIR filters for WL-MMSE estimation using (2) or (3).

III. Iterative WL Estimation

Defining $\tilde{\boldsymbol{z}} = \hat{\tilde{\boldsymbol{\theta}}} - \tilde{\boldsymbol{\theta}}$, the objective function can be stated as follows:

$$\tilde{J}_{opt} = \mathcal{E}\left\{\mathrm{Tr}\left\{\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^T\right\}\right\} \tag{7}$$

$$= \tilde{\mathbf{H}}\mathbf{C}_{\tilde{y}\tilde{y}}\tilde{\mathbf{H}}^T - \tilde{\mathbf{H}}\mathbf{C}_{\tilde{y}\tilde{\theta}} - \mathbf{C}_{\tilde{\theta}\tilde{y}}\tilde{\mathbf{H}}^T + \mathbf{C}_{\tilde{\theta}\tilde{\theta}} \tag{8}$$

where $\mathbf{C}_{\tilde{y}\tilde{y}} = \mathcal{E}\left\{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}^T\right\}$, $\mathbf{C}_{\tilde{\theta}\tilde{y}} = \mathcal{E}\left\{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{y}}^T\right\}$ and $\mathbf{C}_{\tilde{\theta}\tilde{\theta}} = \mathcal{E}\left\{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T\right\}$. For the minimization of MSE in (8), standard rules for differentiation w.r.t the matrix $\tilde{\mathbf{H}} \in \mathbb{R}^{2N\times 2N}$ are applied that lead to the following condition:

$$\tilde{\mathbf{H}}\mathbf{C}_{\tilde{y}\tilde{y}} = \mathbf{C}_{\tilde{\theta}\tilde{y}}. \tag{9}$$

Solving (9) for the Wiener filter matrix $\tilde{\mathbf{H}} = \mathbf{C}_{\tilde{\theta}\tilde{y}}\mathbf{C}^{-1}_{\tilde{y}\tilde{y}}$ requires computationally expensive matrix inverse operation. Instead, we note that (9) is equivalent to solving a system of linear equations $\mathbf{C}_{\tilde{y}\tilde{y}}\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{y}}$ followed by projection onto $\mathbf{C}_{\tilde{\theta}\tilde{y}}$. The conjugate gradient (CG) method solves for $\tilde{\boldsymbol{w}}$ by arriving at the optimal solution step-by-step. It does so by successively looking in mutually conjugate directions in Krylov subspace. The complete algorithm is listed in Table I.

It clear from Table I that, in each iteration, the computational complexity of this IWL algorithm is dominated by steps 3 and 5. For instance, $\mathcal{O}(N^2)$ operations are needed to execute matrix-vector multiplication in step 5. One possible solution is to approximate $\mathbf{C}_{\tilde{y}\tilde{y}}$ by a banded approximate having bandwidth Q. As a consequence, the complexity is reduced to $\mathcal{O}(QN)$ at the cost of approximation error.

The IWL algorithm requires a maximum of N iterations to converge to the optimum solution. Because phase noise has low signal dimensions, a relatively small number of iteration were required to converge to the exact estimate in our simulation environment.

IV. Reduced Dimension WL-MMSE Filtering

As the number of unknown but correlated phase noise terms $\theta(n)$ equals the dimension of $\boldsymbol{y}$, the aim here is to find a reduced-rank realization of $N \times 1$ vector $\boldsymbol{\theta}$ to make the estimation procedure efficient and feasible. Given the observation reference $\bar{\boldsymbol{y}} = \left[\boldsymbol{y}^T\boldsymbol{y}^H\right]^T$, the best linear estimator of $\boldsymbol{\theta}$ in MSE sense is the well-known Wiener filter $\bar{\mathbf{F}} = \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}^{-1}_{\bar{y}\bar{y}}$, where $\mathbf{C}_{\bar{\theta}\bar{y}} = \mathcal{E}\left\{\bar{\boldsymbol{\theta}}\bar{\boldsymbol{y}}^H\right\}$ and $\mathbf{C}_{\bar{y}\bar{y}} = \mathcal{E}\left\{\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}^H\right\}$. Let $\hat{\bar{\boldsymbol{\theta}}}_\eta = \check{\mathbf{F}}\bar{\boldsymbol{y}}$ be a reduced dimensional estimate of $\bar{\boldsymbol{\theta}}$. To find $\check{\mathbf{F}}$, we will consider two algorithms for rank reduction namely Data subspace reduction (DSR) and Parameter subspace reduction (PSR).

A. Data Subspace Reduction via Principal Components

One way of achieving dimensionality reduction is to curb the rank of the observation covariance matrix $\mathbf{C}_{\bar{y}\bar{y}}$. This will essentially project the received vector onto an estimate of lower dimensional signal subspace with significant energy. A reduction in rank follows when N is larger than the signal subspace and in fact, for the phase noise estimator, the energy associated with the parameter subspace is expected to be smaller than the data signal subspace. The DSR estimation is based on the eigen-decomposition:

$$\mathbf{C}_{\bar{y}\bar{y}} = \mathbf{V}_N\boldsymbol{\Lambda}_N\mathbf{V}_N^H \tag{10}$$

Initialize: $\widehat{\boldsymbol{w}}_0 = \mathbf{0}$, $\boldsymbol{p}_0 = -\boldsymbol{g}_0 = \tilde{\boldsymbol{y}}$
for $k = 0$ to $K-1$
$\quad \alpha_{k+1} = \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{p}_k^T \mathbf{C}_{\tilde{y}\tilde{y}} \boldsymbol{p}_k}$
$\quad \widehat{\boldsymbol{w}}_{k+1} = \widehat{\boldsymbol{w}}_k + \alpha_{k+1}\boldsymbol{p}_k$
$\quad \boldsymbol{g}_{k+1} = \boldsymbol{g}_k + \alpha_{k+1}\mathbf{C}_{\tilde{y}\tilde{y}}\boldsymbol{p}_k$
$\quad \beta_{k+1} = \frac{\boldsymbol{g}_{k+1}^T \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^T \boldsymbol{g}_k}$
$\quad \boldsymbol{p}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1}\boldsymbol{p}_k$
end for
$\hat{\bar{\boldsymbol{\theta}}} = \mathbf{C}_{\tilde{\theta}\tilde{y}} \widehat{\boldsymbol{w}}_{K-1}$

TABLE I. THE PROPOSED ITERATIVE WL (IWL) ALGORITHM

where $\mathbf{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{2N}]$ is the orthonormal matrix containing eigenvectors $\boldsymbol{v}_i$ and $\boldsymbol{\Lambda}_N$ is the diagonal matrix with eigen-value ordering $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2N}$. Let $\mathbf{V}_\eta = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{2\eta}]^T$ be a $2N \times 2\eta$ matrix whose columns form the orthonormal basis for the η-dimensional subspace for the received vector $\bar{\boldsymbol{y}}_\eta = \mathbf{V}_\eta \mathbf{V}_\eta^H \bar{\boldsymbol{y}}$ and $\mathbf{C}_{\bar{y}\bar{y}}^{(\eta)^{-1}} := \mathbf{V}_\eta \boldsymbol{\Lambda}_\eta^{-1} \mathbf{V}_\eta^H$. Note that this procedure is dissimilar to [8] where effectively the correlation matrix of the estimand is simply replaced by its low-rank version.

The difference in MSE due to DSR relative to $\hat{\bar{\boldsymbol{\theta}}}$ and the total MSE can be computed as:

$$\varepsilon_\eta^2 = \frac{1}{2}\mathrm{Tr}\left\{\mathcal{E}\left\{\left(\hat{\bar{\boldsymbol{\theta}}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)\left(\hat{\bar{\boldsymbol{\theta}}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)^H\right\}\right\} \tag{11}$$

$$= \frac{1}{2}\mathrm{Tr}\Bigg\{\left(\mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{(\eta)^{-1}} - \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{-1}\right)\mathbf{C}_{\bar{y}\bar{y}} \times \left(\mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{(\eta)^{-1}} - \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{-1}\right)^H\Bigg\} \tag{12}$$

$$= \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}_{\bar{\theta}\bar{y}}\left(\sum_{i=2\eta+1}^{2N}\boldsymbol{v}_i\lambda_i^{-1}\boldsymbol{v}_i^H\right)\mathbf{C}_{\bar{\theta}\bar{y}}^H\right\} \tag{13}$$

$$\mathrm{MSE}\left(\hat{\bar{\boldsymbol{\theta}}}_\eta\right) = \frac{1}{2}\mathrm{Tr}\left\{\mathcal{E}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)^H\right\}\right\} \tag{14}$$

$$= \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}_{\bar{\theta}\bar{\theta}} - \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{(\eta)^{-1}}\mathbf{C}_{\bar{\theta}\bar{y}}^H\right\} \tag{15}$$

where it is obvious that the instability of $\mathbf{C}_{\bar{y}\bar{y}}^{-1}$ computation for low-dimensional $\bar{\boldsymbol{y}}$ has been avoided.[2] Moreover, the estimation error monotonically increases with smaller η as revealed by (15).

For the best performance, one needs to predict the desired signal subspace in advance so as to minimize the effect of summation term in (13). Fortunately, rank reduction can be applied without exact knowledge of the actual signal subspace dimension η due to wider additive noise subspace. The major computational effort is finding eigen-decomposition in (10) requiring $\mathcal{O}\left(N^3\right)$ flops.

B. Parameter Subspace Reduction via GKLT

This method instead reduces the dimension of the range space of the filter $\breve{\mathbf{F}}$ (column space of $\bar{\mathbf{F}}$) that restricts $\hat{\bar{\boldsymbol{\theta}}}$ to lie in a reduced dimension space. The objective function to minimize the covariance of error: $\bar{\boldsymbol{z}} = \bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}_\eta$ i.e., $\bar{J}_{opt}^{(\eta)} = \mathcal{E}\left\{\mathrm{Tr}\left\{\bar{\boldsymbol{z}}\bar{\boldsymbol{z}}^T\right\}\right\}$, which is known to have GKLT solution [10] explained in the following.

The key step here is to find the rank-constrained filter matrix $\breve{\mathbf{F}} \in \mathbb{C}^{2N \times 2N}$ while minimizing the trace of the extra covariance (or model bias) :

$$\bar{J}_{opt}^{(\eta)} - \bar{J}_{opt}^{(N)} = \mathcal{E}\left\{\mathrm{Tr}\left\{\left(\breve{\mathbf{F}}\bar{\boldsymbol{y}} - \bar{\mathbf{F}}\bar{\boldsymbol{y}}\right)\left(\breve{\mathbf{F}}\bar{\boldsymbol{y}} - \bar{\mathbf{F}}\bar{\boldsymbol{y}}\right)^H\right\}\right\}$$
$$= \mathrm{Tr}\left\{\left(\breve{\mathbf{F}} - \bar{\mathbf{F}}\right)\mathbf{C}_{\bar{y}\bar{y}}\left(\breve{\mathbf{F}} - \bar{\mathbf{F}}\right)^H\right\}. \tag{16}$$

Now let us decompose the positive semi-definite matrix $\mathbf{C}_{\bar{y}\bar{y}} = \mathbf{C}_{\bar{y}\bar{y}}^{1/2}\mathbf{C}_{\bar{y}\bar{y}}^{1/2^H}$, then it is obvious that $\varepsilon_\eta^2 = \bar{J}_{opt}^{(\eta)} - \bar{J}_{opt}^{(N)}$ is:

$$\varepsilon_\eta^2 = \left\|\left(\breve{\mathbf{F}} - \bar{\mathbf{F}}\right)\mathbf{C}_{\bar{y}\bar{y}}^{1/2}\right\|_F^2. \tag{17}$$

In a result known as Eckart-Young-Mirsky Theorem [15], this corresponds to finding a truncated SVD:

$$\breve{\mathbf{F}}\mathbf{C}_{\bar{y}\bar{y}}^{1/2} = \sum_{i=1}^{2\eta}\boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^H \tag{18}$$

$$\text{and} \quad \breve{\mathbf{F}} = \sum_{i=1}^{2\eta}\boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^H\mathbf{C}_{\bar{y}\bar{y}}^{-1/2} \tag{19}$$

where the SVD of $\mathbf{C}_{WL} = \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{-1/2} = \mathbf{U}_N\boldsymbol{\Sigma}_N\mathbf{V}_N^H = \sum_{i=1}^{2N}\boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^H$. Due to the fact that improperness destroys even multiplicity of the singular-values[3], (19) approximates $2 \times \eta$ significant canonical correlations in the ordering $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{2\eta}$.

Proposition 1: The estimation MSE $\left(\hat{\bar{\boldsymbol{\theta}}}_\eta\right)$ between $\bar{\boldsymbol{\theta}}$ and its low-rank approximate $\hat{\bar{\boldsymbol{\theta}}}_\eta$ is the sum of model bias ε_η^2 given by:

$$\varepsilon_\eta^2 = \frac{1}{2}\sum_{i=2\eta+1}^{2N}\sigma_i^2 \tag{20}$$

and the error variance of the full-rank estimator $\nu^2 = \frac{1}{2}\mathcal{E}\left\{\mathrm{Tr}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)^H\right\}\right\}$ known to be:

$$\nu^2 = \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}_{\bar{\theta}\bar{\theta}} - \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{-1}\mathbf{C}_{\bar{y}\bar{\theta}}\right\} \tag{21}$$

$$= \frac{1}{2}\sum_{i=1}^{2N}\left(\sigma_{\bar{\theta},i}^2 - \sigma_i^2\right) \tag{22}$$

where $\sigma_{\bar{\theta},i}^2$ are the singular-values of $\mathbf{C}_{\bar{\theta}\bar{\theta}}$.

Proof: See appendix A. ■

Remark 1: A crucial implication of (30) is that reducing dimension η monotonically increases the error variance e_η^2 by elevating the model bias ε_η^2 or equivalently by shifting purged singular-values from the first to second summation in (30).

Remark 2: If we neglect ν^2 for the moment and assume that in SL case $\mathbf{C}_{SL} = \mathbf{C}_{\bar{\theta}\bar{y}}\mathbf{C}_{\bar{y}\bar{y}}^{-1/2}$ has replicated singular-values $\vec{\sigma}_1, \vec{\sigma}_2, \ldots, \vec{\sigma}_N$, then the reduced-rank WL system curbs model bias by an amount:

$$\delta_\varepsilon^2 = 2\sum_{i=\eta+1}^{N}\vec{\sigma}_i^2 - \sum_{i=2\eta+1}^{2N}\sigma_i^2 \geq 0 \tag{23}$$

where the inequality results from singular-value analysis for matrix perturbations [15]. Interestingly, if one considers that the distribution of $\bar{\boldsymbol{\theta}}$ (e.g., for models in [1], [2], [5], [16])

[2] There is a non-zero probability of any $x(n)$ being zero, if its discrete Fourier transform $X(m)$ has spectral nulls implying that $\mathrm{diag}\left[\boldsymbol{x}\right]$ can be rank-deficient.

[3] Proper case has even-multiplicity i.e., $\sigma_{2i-1} = \sigma_{2i}$, $\forall\, i = 1, 2, \cdots, N$.

effects the composition of $\mathbf{C}_{\tilde{\theta}\tilde{\theta}}$ and hence, the imbalance between σ_{2i-1} and σ_{2i}, the gap in (23) with decreasing η values can not only widen but in some cases close down due to spectrum dependent disparity in the singular values σ_i.

The major computational burden in (19) is the computation of the inverse-square-root matrix $\mathbf{C}_{\tilde{y}\tilde{y}}^{-1/2}$ requiring $\mathcal{O}\left(N^3\right)$. To circumvent the explicit computation, an implicit technique is the fixed-rank matrix factorization (FMF). Let $\check{\mathbf{H}} = \left(\tilde{\mathbf{U}}_\eta \tilde{\boldsymbol{\Sigma}}_\eta^{1/2}\right)\left(\tilde{\boldsymbol{\Sigma}}_\eta^{1/2} \tilde{\mathbf{V}}_\eta^T\right) = \mathbf{Q}\mathbf{R}^T$, where $\mathbf{Q} \in \mathbb{R}_*^{2N\times 2\eta}$, $\mathbf{R} \in \mathbb{R}_*^{2N\times 2\eta}$ and $\mathbb{R}_*^{n\times m}$ is the set of fixed column rank $n \times m$ real matrices. The solution $\hat{\tilde{\boldsymbol{\theta}}}_\eta = \check{\mathbf{H}}\tilde{\boldsymbol{y}}$ is obtained from an MSE minimization problem:

$$\tilde{J}_{opt}^{(\eta)} = \mathcal{E}\left\{\mathrm{Tr}\left\{\left(\check{\mathbf{H}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{\theta}}\right)\left(\check{\mathbf{H}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{\theta}}\right)^T\right\}\right\} \tag{24}$$

$$= \mathrm{Tr}\left\{\mathbf{C}_{\tilde{\theta}\tilde{\theta}} + \mathbf{Q}\mathbf{R}^T\mathbf{C}_{\tilde{y}\tilde{y}}\mathbf{R}\mathbf{Q}^T - \mathbf{Q}\mathbf{R}^T\mathbf{C}_{\tilde{y}\tilde{\theta}} - \mathbf{C}_{\tilde{\theta}\tilde{y}}\mathbf{R}\mathbf{Q}^T\right\} \tag{25}$$

where the optimal solution satisfies following conditions at the stationary point:

$$-\mathbf{C}_{\tilde{\theta}\tilde{y}}\mathbf{R} + \mathbf{Q}\mathbf{R}^T\mathbf{C}_{\tilde{y}\tilde{y}}\mathbf{R} = \mathbf{0} \tag{26}$$

$$-\mathbf{C}_{\tilde{y}\tilde{\theta}}^T\mathbf{Q} + \mathbf{C}_{\tilde{y}\tilde{y}}\mathbf{R}\mathbf{Q}^T\mathbf{Q} = \mathbf{0}. \tag{27}$$

Due to the inter-dependence between optimization variables, closed-form results are rather difficult to derive. However, a useful technique is to minimize $\tilde{J}_{opt}^{(\eta)}$ w.r.t $\mathbf{Q}$ and $\mathbf{R}$ by alternative optimization. At the k-th iteration, parameter updates comprises of:

$$\mathbf{Q}(k+1) = \mathbf{C}_{\tilde{\theta}\tilde{y}}\mathbf{R}(k)\left(\mathbf{R}^T(k)\mathbf{C}_{\tilde{y}\tilde{y}}\mathbf{R}(k)\right)^{-1} \tag{28}$$

$$\mathbf{R}(k+1) = \mathbf{C}_{\tilde{y}\tilde{y}}^{-1}\mathbf{C}_{\tilde{\theta}\tilde{y}}\mathbf{Q}^\dagger(k+1). \tag{29}$$

For most practical oscillators, $\eta \ll N$ implying that $\mathbf{Q}$ and $\mathbf{R}$ are tall matrices and the algorithm can be realized with $\mathcal{O}\left(N^2\eta\right)$ computational operations at each iteration.

V. Test Scenarios with Numerical Results

Numerical results are presented in this section with two goals. First, the performance improvement of WL over SL is quantified under various conditions. Next, the effectiveness of the proposed algorithms are tested in OFDM systems with transmission taking place over multi-tap Rayleigh fading channels (except Fig. 1) which is equalized by a single-tap LMMSE equalizer. Simulation environment for all numerical experiments consists of K-sized FFT with $K = N = 64$ and the total bandwidth occupation of 10 MHz. Perfect decisions are assumed except in Fig. 4 where estimators operate in decision-directed manner using convolutional error codes and soft-decisions. Two phase noise models are used: In Model-A, $\phi(n)$ is a Wiener-Lévy process [1], [5] with the distribution of increments: $\phi(n) - \phi(n-1) \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$, where as Model-B simulates PSD of phase noise with a linear-decay behavior (refer to [16] with parameters: $a = 7$, $b = 2$, $c = 11$, $f_l = 10$ KHz and $f_h = 1$ MHz).

Fig. 1 plots the MSE gain of WL estimator: $\Delta\mathrm{MSE} = \frac{1}{N}\mathcal{E}_x\left\{\sum_{n=1}^{N}\frac{\mathrm{MSE}\left(\hat{\theta}'_{\mathrm{SL}}(n)\right)}{\mathrm{MSE}\left(\hat{\theta}'_{\mathrm{WL}}(n)\right)}\right\}$ for PSR and DSR methods. As mentioned earlier, the gain varies with rank for both phase noise models but depicts a unique maxima for Model-B. However, the maximum gain achieving rank is independent of SNR and the gain increases with higher SNR. In general, all WL estimators converge to full-rank gain that is found to be lower than 3-dB as expected from [11].

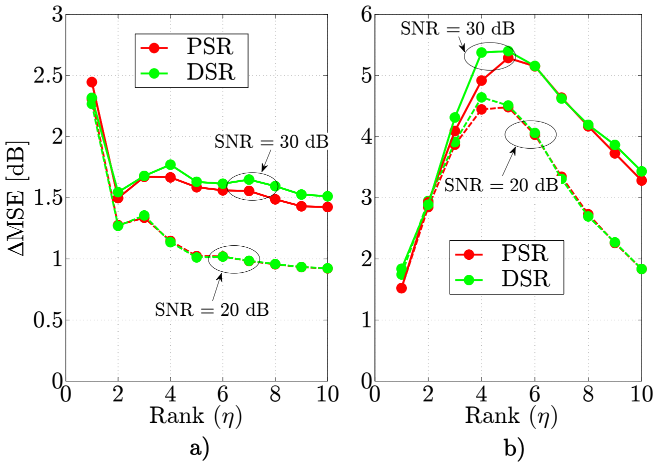

Fig. 1. Gain of WL w.r.t SL system in terms of MSE versus estimator rank over AWGN channel. a) Model-A ($\sigma_\zeta^2 = 10^{-4}$) and b) Model-B.

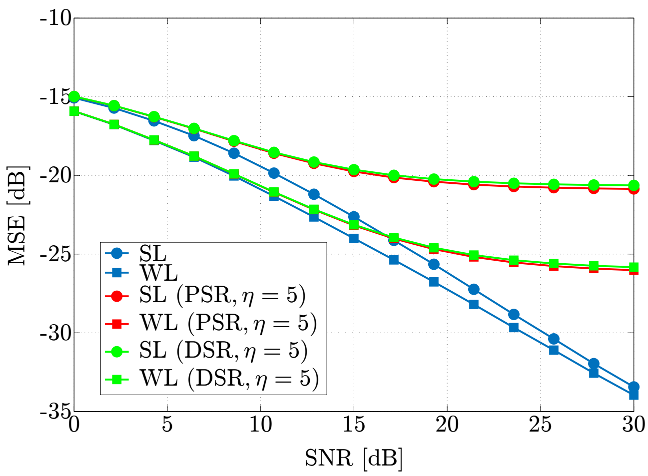

Fig. 2. MSE of phase estimates versus SNR with phase noise Model-B.

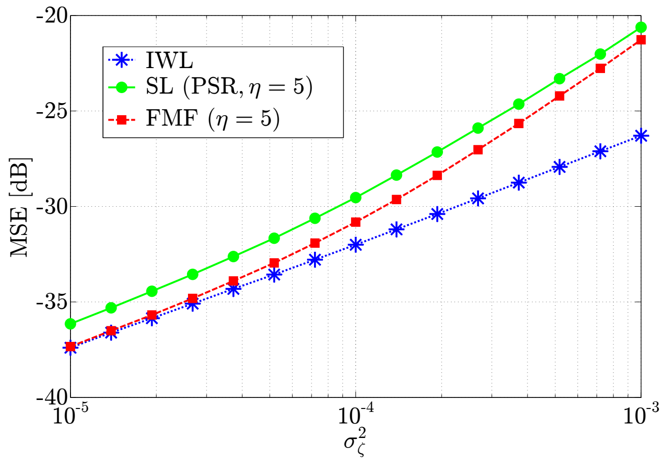

Fig. 3. MSE of phase estimates versus Wiener phase increment variance σ_ζ^2 for SNR= 20 dB.

In Fig. 2, we compare the sum-MSE of the SL and WL estimators. In terms of MSE, we do not find much difference in the performance of PSR and DSR methods. When the full-rank WL estimator has little to offer ($1 - 2$ dB at most), its reduced-rank version improves the MSE performance up to 5-dB w.r.t SL counterparts.

The reduction of estimation precision with higher phase noise severity is demonstrated in Fig. 3 for the PSR method. It is obvious that low-rank version converges to full-rank WL estimator for lower σ_ζ^2 due to small dimensionality of $\boldsymbol{\theta}$ where

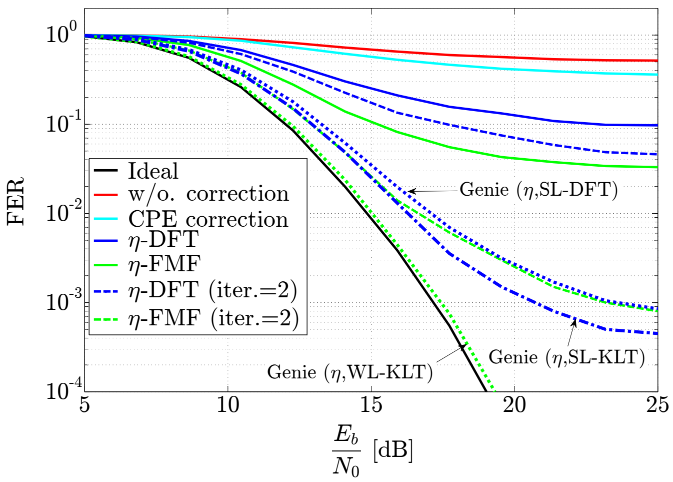

Fig. 4. Frame error rate (FER) performance of phase noise compensation schemes with $\eta = 5$ and Model-B phase noise. A codeword spans 3 OFDM symbols using rate-$1/2$ convolutional code and mapped to 64-QAM symbols in a bit-interleaved-coded-modulation (BICM) structure with $802.11a/g$ like subcarrier mapping. The CPE is estimated using pilot tones whereas CSI is known perfectly.

as non-optimal rank selection in higher phase noise variance σ_ζ^2 conditions reduces gap between η-rank WL and SL estimators. In our extensive simulations, residual error in IWL persistently dropped below -50 dB in 7 iterations, whereas it took FMF 15 iterations to achieve the same feat due to linear convergence.

Finally, we put our proposed method to test in Fig. 4 for a cyclic-prefixed OFDM system and analyze FER performance against common-phase-error (CPE) correction method [6] and the state-of-the-art DFT-compressed (η-DFT) estimator in SL MMSE class [1] with $2u + 1 = \eta$, $l_P = N$ and consistent compensation technique. We can observe that the proposed method η-FMF outperforms η-DFT at FER of 10^{-1} by 7-dB without iterations and η-FMF without iteration is better than η-DFT with 2 iterations. Both techniques have comparable rank definition i.e., Karhunen-Loeve transform (KLT) vs Fourier transform basis functions in addition to the fact that their complexities are scaled by the estimation order. One plausible reason of the superior results for η-FMF, in addition to WLE gain (Fig. 1), is the higher degree of component compression (around 2-dB better SNR) achieved by KLT than by any DFT-based method as obvious from the genie plots in Fig. 4.

VI. Conclusions

Optimal and reduced-rank widely-linear filters for phase noise estimation have been presented based on MMSE criterion. For large filter lengths, efficient implementations were proposed using iterative methods and shown to allow reasonable complexity reduction relative to explicit computations. Numerical results indicate that by properly selecting estimation-rank, the complexity-reduced method can approach the precision of a full-rank estimator, whereas performance gains w.r.t its strictly linear counterparts are substantial, both from the perspective of estimation accuracy and the digital baseband error performance.

Appendix A
Proof Of Proposition 1

Firstly, we revisit the orthogonality property of the full-rank estimate and its error i.e., $\mathcal{E}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)\hat{\bar{\boldsymbol{\theta}}}^H\right\} = \mathbf{0}$. It is important to note that (19) places $\hat{\bar{\boldsymbol{\theta}}}_\eta$ in a reduced dimension subset of subspace spanned by $\hat{\bar{\boldsymbol{\theta}}}$ implying that the former can be obtained by a projection operator as $\hat{\bar{\boldsymbol{\theta}}}_\eta = \mathbf{U}_\eta \mathbf{U}_\eta^H \hat{\bar{\boldsymbol{\theta}}} = \mathbf{P}_\eta \hat{\bar{\boldsymbol{\theta}}}$, where $\mathbf{U}_\eta$ is a tall matrix with orthonormal columns and $\mathbf{P}_\eta$ is the projection on the column space of $\mathbf{U}_\eta$. This in turn means that $\mathcal{E}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)\hat{\bar{\boldsymbol{\theta}}}_\eta^H\right\} = \mathcal{E}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)\hat{\bar{\boldsymbol{\theta}}}^H\right\}\mathbf{P}_\eta = \mathbf{0}$ also holds. Finally, the error variance of the estimate $\hat{\bar{\boldsymbol{\theta}}}_\eta$ can be derived as:

$$\begin{aligned}
\mathrm{MSE}\left(\hat{\bar{\boldsymbol{\theta}}}_\eta\right) &= \frac{1}{2}\mathcal{E}\left\{\mathrm{Tr}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)^H\right\}\right\} \\
&= \frac{1}{2}\mathcal{E}\left\{\mathrm{Tr}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}} + \hat{\bar{\boldsymbol{\theta}}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}} + \hat{\bar{\boldsymbol{\theta}}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)^H\right\}\right\} \\
&= \frac{1}{2}\mathcal{E}\left\{\mathrm{Tr}\left\{\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)\left(\bar{\boldsymbol{\theta}} - \hat{\bar{\boldsymbol{\theta}}}\right)^H + \left(\hat{\bar{\boldsymbol{\theta}}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)\left(\hat{\bar{\boldsymbol{\theta}}} - \hat{\bar{\boldsymbol{\theta}}}_\eta\right)^H\right\}\right\} \\
&= \frac{1}{2}\left(\sum_{i=1}^{2\eta}\left(\sigma_{\bar{\theta},i}^2 - \sigma_i^2\right) + \sum_{i=2\eta+1}^{2N}\sigma_{\bar{\theta},i}^2\right) \qquad (30)
\end{aligned}$$

that shows the estimation error consists of two independent components: the estimation error variance and the purged dimension variance.

References

[1] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *Communications, IEEE Transactions on*, vol. 55, no. 8, pp. 1607–1616, Aug 2007.

[2] A. Ishaque and G. Ascheid, "Efficient MAP-based estimation and compensation of phase noise in MIMO-OFDM receivers," *Elsevier AEÜ- International Journal of Electronics and Communications*, 2013.

[3] K. Nikitopoulos, S. Stefanatos, and A. Katsaggelos, "Decision-aided compensation of severe phase-impairment-induced inter-carrier interference in frequency-selective OFDM," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 4, pp. 1614–1619, April 2009.

[4] R. Casas, S. Biracree, and A. Youtz, "Time domain phase noise correction for OFDM signals," *Broadcasting, IEEE Transactions on*, vol. 48, no. 3, pp. 230–236, Sep 2002.

[5] H. Mehrpouyan, A. Nasir, S. Blostein, T. Eriksson, G. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4790–4807, Sept 2012.

[6] S. Wu, P. Liu, and Y. Bar-Ness, "Phase noise estimation and mitigation for OFDM systems," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 12, pp. 3616–3625, December 2006.

[7] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862 –1874, May 2007.

[8] R. Wang, H. Mehrpouyan, M. Tao, and Y. Hua, "Channel estimation, carrier recovery, and data detection in the presence of phase noise in OFDM relay systems," *CoRR*, vol. abs/1309.7564, 2013. [Online]. Available: http://arxiv.org/abs/1309.7564

[9] X. Wang and H. Poor, "Blind multiuser detection: a subspace approach," *Information Theory, IEEE Transactions on*, vol. 44, no. 2, Mar 1998.

[10] Y. Hua and W. Liu, "Generalized karhunen-loeve transform," *Signal Processing Letters, IEEE*, vol. 5, no. 6, pp. 141–142, June 1998.

[11] P. Schreier, L. Scharf, and C. Mullis, "Detection and estimation of improper complex random signals," *Information Theory, IEEE Transactions on*, vol. 51, no. 1, pp. 306–312, 2005.

[12] D. Mattera, L. Paura, and F. Sterle, "Widely linear decision-feedback equalizer for time-dispersive linear MIMO channels," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2525–2536, July 2005.

[13] D. Darsena, G. Gelli, L. Paura, and F. Verde, "Widely linear equalization and blind channel identification for interference-contaminated multicarrier systems," *Signal Processing, IEEE Transactions on*, vol. 53, no. 3, pp. 1163–1177, March 2005.

[14] C. Cheong Took, S. Douglas, and D. Mandic, "On approximate diagonalization of correlation matrices in widely linear signal processing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 3, March 2012.

[15] L. Mirsky, "Symmetric gauge functions and unitarily invariant norms," *Quart. J. Math. Oxford*, vol. 11, pp. 50–59, 1960.

[16] Y.-C. Liao and K.-C. Chen, "Estimation of stationary phase noise by the autocorrelation of the ICI weighting function in OFDM systems," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 12, pp. 3370–3374, December 2006.