Augmented Performance Bounds on Strictly Linear and Widely Linear Estimators With Complex Data

Yili Xia, *Member, IEEE*, and Danilo P. Mandic, *Fellow, IEEE*

***Abstract*—Novel physical insights are provided into the performances of strictly linear (SL) and widely linear (WL) estimators of the generality of complex-valued data, both proper (second order circular) and improper (second-order noncircular). This is achieved by first performing a novel complementary mean square error (CMSE) analysis, in order to quantify the degrees of improperness (second order noncircularity) of the SL and WL estimation errors. The exact bounds on the CMSE difference between the SL and WL estimators are investigated to show that only a joint consideration of the standard MSE analysis and the proposed CMSE analysis has enough degrees of freedom for a rigorous account of the performance of WL and SL estimators. This also makes it possible to rigourously quantify the contributions to the WL performance advantage from the individual real and imaginary channels, an important finding not possible to obtain by using the standard MSE analysis only.**

***Index Terms*—Widely linear model, augmented complex statistics, improperness (second order noncircularity), complementary mean square error (CMSE), real-imaginary analysis.**

I. Introduction

RANDOM signals in the complex domain $\mathbb{C}$ arise in many areas of science and engineering, and their accurate and robust estimation is thus of fundamental interest. Standard statistical signal processing approaches typically impose a restrictive assumption on a complex-valued zero-mean random vector, $\mathbf{x} \in \mathbb{C}^{N\times 1}$, in the form of *properness* or *second order circularity* [1], [2]. In practical terms, a circular random vector has a rotation-invariant probability distribution, while a proper (second order circular) vector $\mathbf{x}$ is uncorrelated with its complex conjugate $\mathbf{x}^*$, or equivalently, its real and imaginary components are uncorrelated, and are with equal powers. The above assumption of properness (second order circularity) is convenient in many respects because it simplifies computations and makes the complex-valued signal processing a straightforward extension of its real-valued counterpart, in the sense that second order statistical properties are described by the covariance matrix only, $\mathbf{R} = E[\mathbf{x}\mathbf{x}^H]$, where $E[\cdot]$ denotes the statistical expectation and $(\cdot)^H$ the Hermitian transpose operator.

Consequently, in classical mean square estimation in $\mathbb{C}$, a scalar random variable (estimandum) y is estimated based on the random vector regressor, $\mathbf{x}$, and the estimate $\hat{y}$ that minimises the mean square error (MSE) is given by the conditional expectation $\hat{y} = E[y|\mathbf{x}]$. When the zero-mean pair $\{y, \mathbf{x}\}$ is jointly circularly distributed Gaussian, the optimal solution is a strictly linear (SL) estimator, which is directly inherited from the real domain $\mathbb{R}$, and is given by [3]

$$\hat{y}_{\mathrm{SL}} = \mathbf{w}^H \mathbf{x}, \tag{1}$$

where $\mathbf{w} \in \mathbb{C}^{N\times 1}$ is a coefficient vector.

However, recent results in the so-called *augmented complex statistics* [4]–[7] show that the covariance matrix, $\mathbf{R}$, cannot completely describe the second order behaviour of general complex-valued signals, and another second order moment, called the complementary covariance (pseudo-covariance) matrix, $\mathbf{P} = E[\mathbf{x}\mathbf{x}^T]$, should also be taken into account, especially when processing second order noncircular (improper) data. The improperness is characterised by a non-zero $\mathbf{P} \neq \mathbf{0}$, and the complementary covariance matrix $\mathbf{P}$ therefore vanishes only when $\mathbf{x}$ is proper. Intuitively speaking, in order for an estimator to utilise all the available second order information in both proper and improper signals, a general estimation framework in $\mathbb{C}$ should depend on both the signals themselves and their complex conjugates. In the context of mean square error estimation in $\mathbb{C}$, this leads to the well-known widely linear (WL) estimator, given by [6]–[8]

$$\hat{y}_{\mathrm{WL}} = \mathbf{h}^H \mathbf{x} + \mathbf{g}^H \mathbf{x}^*, \tag{2}$$

where $\mathbf{h}$ and $\mathbf{g}$ are the $N \times 1$ complex-valued coefficient vectors associated with $\mathbf{x}$ and $\mathbf{x}^*$, respectively.

A rigourous MSE analysis of both SL and WL estimators with complex-valued Gaussian data was conducted in a seminal work in [8], which showed that apart from a special case for which the WL estimator in (2) reduces to the SL estimator in (1), both giving identical MSE performance, the WL estimator offers significant performance gains over the SL one. This finding has spurred extensive use of WL estimation in numerous applications, including signal processing and communications [9]–[29], where improper signals occur due to their underlying signal generating physics, so that the traditional SL processing framework is inadequate due to the inherent undermodelling. For instance, in radio communications, the received

Manuscript received June 18, 2017; revised October 24, 2017; accepted November 1, 2017. Date of publication November 13, 2017; date of current version December 11, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Marco Lops. This work was supported in part by the National Natural Science Foundation of China under Grants 61401094 and 61771124, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20140645, and in part by the Fundamental Research Funds for the Central Universities (Grant 2242016K41050).

(Corresponding author: Yili Xia.)

Y. Xia is with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: yili_xia@seu.edu.cn).

D. P. Mandic is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: d.mandic@imperial.ac.uk).

Digital Object Identifier 10.1109/TSP.2017.2773428

baseband signals over fading channels exhibit second order noncircular (improper) behaviours when the transmission system adopts noncircular real-valued and complex-valued constellation schemes, e.g., binary phase shift-keying (BPSK), 8 phase shift-keying (8PSK), and offset quadrature amplitude modulation (OQAM) [18]–[22]. In areas such as radio communications, spectral sensing, beam-forming, noncircularity also occurs in the tramsmitted/received signal due to the amplitude and phase imbalances between its in-phase (I) and quadrature (Q) components, even when second order circular modulation schemes are employed [23]–[29]. In addition, the use of improper signals and augmented complex statistics has been reported in power systems [30]–[32], biomedical engineering [33], and renewable energy [34], [35]. Despite these advances, it is important to note that, from the perspective of augmented complex statistics, the current quantification of performance advantages of WL estimators over their SL counterparts still provides limited physical insights into both estimators, since the existing approaches implicitly omit the complementary second order statistics of the estimation errors - a key feature of improper data.

To address this issue, we first propose a novel complementary mean square error (CMSE) analysis of both the SL and WL estimators, in order to quantify the degrees of improperness of their respective estimation errors, a critical parameter for improper data. Furthermore, drawing from the physics of impropriety, which states that the complementary variance of a complex-valued zero-mean random variable x, that is, $\widetilde{\sigma}_x^2 = E[x^2]$, is bounded by the standard variance (power), $\sigma_x^2 = E[xx^*]$, that is $|\widetilde{\sigma}_x^2| \leq \sigma_x^2$ [6], [7], [36], the boundedness of the CMSE difference between the SL and WL estimators is rigorously and comprehensively examined. In this way, we show that only a joint consideration of the standard MSE analysis and the proposed CMSE analysis makes it possible to examine the complete second order performance of both the estimators. This also enables more physical insight into the underlying physics of the contribution of the real and imaginary channels of both the estimators, and hence, the precise quantification of the advantages of an WL estimator over its SL counterpart. This important aspect is still missing in the literature and, in general, owing to the insufficient degrees of freedom, cannot be examined by using the standard MSE analysis in [8].

II. Standard MSE Analysis of WL and SL Estimators

This section briefly summarises the standard MSE analysis of both WL and SL estimators, conducted in [8].

The aim of the WL estimator is to find the optimal weight vectors $\mathbf{h}$ and $\mathbf{g}$ in (2) which minimise the MSE, denoted by $E[|e_{\mathrm{WL}}|^2]$, where

$$e_{\mathrm{WL}} = y - \hat{y}_{\mathrm{WL}} \tag{3}$$

is the estimation error. For this purpose, the first point to note is that the set of scalar complex random variables z, spanned by $z = \mathbf{a}^H\mathbf{x} + \mathbf{b}^H\mathbf{x}^*$, where $\mathbf{a}$ and $\mathbf{b}$ belong to $\mathbb{C}^{N\times 1}$, constitutes a linear space. It becomes a Hilbert subspace with the scalar product $\langle z_1, z_2\rangle = E[z_1 z_2^*]$. As a result, $\hat{y}_{\mathrm{WL}}$ represents a projection of y onto this subspace, while the associated estimation error e_{WL} is characterised by the orthogonality principle, where

$$e_{\mathrm{WL}} \perp \mathbf{x}, \tag{4}$$

and

$$e_{\mathrm{WL}} \perp \mathbf{x}^*. \tag{5}$$

The symbol $\perp$ indicates that all the components of $\mathbf{x}$ and $\mathbf{x}^*$ are orthogonal to e_{WL} (their scalar product is zero). From (3), the first orthogonality condition in (4) can be written in terms of the statistical expectation, to yield[1]

$$E[y^*\mathbf{x}] = E[\hat{y}_{\mathrm{WL}}^*\mathbf{x}]. \tag{6}$$

In a similar way, by considering (3) and (5), we have

$$E[y^*\mathbf{x}^*] = E[\hat{y}_{\mathrm{WL}}^*\mathbf{x}^*]. \tag{7}$$

Upon replacing $\hat{y}_{\mathrm{WL}}$ in (6) and (7) with the expression in (2), we obtain

$$\mathbf{R}\mathbf{h} + \mathbf{P}\mathbf{g} = \mathbf{r}, \tag{8}$$

$$\mathbf{P}^*\mathbf{h} + \mathbf{R}^*\mathbf{g} = \mathbf{s}^*, \tag{9}$$

where

$$\mathbf{r} = E[y^*\mathbf{x}],\ \mathbf{s} = E[y\mathbf{x}]. \tag{10}$$

After a few mathematical manipulations and considering that the covariance matrix $\mathbf{R}$ is Hermitian while the complementary covariance matrix $\mathbf{P}$ is symmetric, from (8) and (9) we obtain the optimal weight vectors, $\mathbf{h}$ and $\mathbf{g}$, in the form [8]

$$\mathbf{h} = [\mathbf{R} - \mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1}[\mathbf{r} - \mathbf{P}\mathbf{R}^{-*}\mathbf{s}^*], \tag{11}$$

$$\mathbf{g} = [\mathbf{R}^* - \mathbf{P}^*\mathbf{R}^{-1}\mathbf{P}]^{-1}[\mathbf{s}^* - \mathbf{P}^*\mathbf{R}^{-1}\mathbf{r}]. \tag{12}$$

where the symbol $\mathbf{R}^{-*}$ designates the operation $(\mathbf{R}^{-1})^*$. Upon combining with (3), (8) and (9), the corresponding MSE of the WL estimator, that is, $E[|e_{\mathrm{WL}}|^2]$, is given by

$$E[|e_{\mathrm{WL}}|^2] = E[|y|^2] - (\mathbf{h}^H\mathbf{r} + \mathbf{g}^H\mathbf{s}^*). \tag{13}$$

The SL estimator in (1) is obviously more restrictive as compared with its WL counterpart, in the sense that its estimate $\hat{y}_{\mathrm{SL}}$ represents a projection of y onto a restricted linear Hilbert subspace defined by only $\mathbf{x}$. In other words, its MSE, denoted as $E[|e_{\mathrm{SL}}|^2]$, and given by

$$e_{\mathrm{SL}} = y - \hat{y}_{\mathrm{SL}}, \tag{14}$$

requires only the single orthogonality condition

$$e_{\mathrm{SL}} \perp \mathbf{x}. \tag{15}$$

From (14) and (15), we can now write

$$E[\hat{y}_{\mathrm{SL}}^*\mathbf{x}] = E[y^*\mathbf{x}] = \mathbf{r}. \tag{16}$$

Substitution of (1) into (16) yields

$$\mathbf{R}\mathbf{w} = \mathbf{r}, \tag{17}$$

[1] By definition, the orthogonality condition $e_{\mathrm{WL}} \perp \mathbf{x}$ leads to $E[e_{\mathrm{WL}}\mathbf{x}^*] = 0$. For mathematical elegance, a complex conjugate operation $(\cdot)^*$ is applied on both sides, after the $E[\cdot]$ operation, which yields (6).

so that

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{r}, \tag{18}$$

and hence, the MSE of the SL estimator, that is, $E[|e_{\mathrm{SL}}|^2]$, can be expressed as

$$E[|e_{\mathrm{SL}}|^2] = E[|y|^2] - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{r}. \tag{19}$$

Upon combining (13) and (19), the performance advantage of the WL estimator over the SL one can be characterised by the difference of the respective MSEs, given by

$$\triangle\mathrm{MSE} = E[|e_{\mathrm{SL}}|^2] - E[|e_{\mathrm{WL}}|^2], \tag{20}$$

and after a few mathematical manipulations, we have [8]

$$\triangle\mathrm{MSE} = [\mathbf{s} - \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*]^H[\mathbf{R} - \mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1}[\mathbf{s} - \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*]. \tag{21}$$

The value of $\triangle$MSE is nonnegative since the matrix $[\mathbf{R} - \mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]$ is positive definite, indicating that the WL estimator always yields an MSE that is smaller than that of the SL estimator when $\mathbf{s} \neq \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$, or at the very least equal to that of the SL estimator when $\mathbf{s} = \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ [8].

Remark 1: The standard MSE analysis implicitly assumes that the estimation errors of both estimators, that is, e_{WL} and e_{SL}, are strictly proper, so that their complementary second order statistics are omitted in (21), although there is no justification to support this assumption.

III. Proposed CMSE Analysis of WL and SL Estimators

To investigate the degrees of improperness of both the WL and SL estimation errors, according to the augmented complex statistics [4]–[7], we next consider the complementary variance of both the estimation errors, defined as $E[e_{\mathrm{WL}}^2]$ and $E[e_{\mathrm{SL}}^2]$. Analogously to the definition of MSE, we refer to these as complementary MSEs (CMSEs) [37], [38]. Note that unlike the real-valued MSE, the CMSE is in general complex-valued.

Upon combining (2), (3) and (10), the CMSE of a WL estimator can be expressed as

$$\begin{aligned}\mathrm{CMSE}_{\mathrm{WL}} &= E[e_{\mathrm{WL}}^2] = E[(y - \hat{y}_{\mathrm{WL}})^2]\\ &= E[y^2] + E[\hat{y}_{\mathrm{WL}}^2] - 2E[y\hat{y}_{\mathrm{WL}}]\\ &= E[y^2] + E[(\mathbf{h}^H\mathbf{x} + \mathbf{g}^H\mathbf{x}^*)^2] - 2(\mathbf{h}^H\mathbf{s} + \mathbf{g}^H\mathbf{r}^*)\\ &= E[y^2] + [\mathbf{h}^H\mathbf{P} + \mathbf{g}^H\mathbf{R}^* - \mathbf{s}^T]\mathbf{h}^*\\ &\quad + [\mathbf{h}^H\mathbf{R} + \mathbf{g}^H\mathbf{P}^* - \mathbf{r}^H]\mathbf{g}^* - (\mathbf{h}^H\mathbf{s} + \mathbf{g}^H\mathbf{r}^*).\end{aligned} \tag{22}$$

According to (11) and (12), the second and third terms on the right hand side (RHS) of (22) vanish, to yield

$$E[e_{\mathrm{WL}}^2] = E[y^2] - (\mathbf{h}^H\mathbf{s} + \mathbf{g}^H\mathbf{r}^*). \tag{23}$$

For the SL estimator, upon using (1), (10), (14), and (18), its CMSE becomes

$$\begin{aligned}\mathrm{CMSE}_{\mathrm{SL}} &= E[e_{\mathrm{SL}}^2] = E[(y - \hat{y}_{\mathrm{SL}})^2]\\ &= E[y^2] + E[\hat{y}_{\mathrm{SL}}^2] - 2E[y\hat{y}_{\mathrm{SL}}]\\ &= E[y^2] + \mathbf{r}^H\mathbf{R}^{-1}\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^* - 2\mathbf{r}^H\mathbf{R}^{-1}\mathbf{s}.\end{aligned} \tag{24}$$

Now, the difference in the CMSE (or degree of improperness) of both the estimators can be quantified as

$$\triangle\mathrm{CMSE} = \mathrm{CMSE}_{\mathrm{SL}} - \mathrm{CMSE}_{\mathrm{WL}} = E[e_{\mathrm{SL}}^2] - E[e_{\mathrm{WL}}^2], \tag{25}$$

and based on (23) and (24), we arrive at

$$\triangle\mathrm{CMSE} = \mathbf{r}^H\mathbf{R}^{-1}\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^* - 2\mathbf{r}^H\mathbf{R}^{-1}\mathbf{s} + \mathbf{h}^H\mathbf{s} + \mathbf{g}^H\mathbf{r}^*. \tag{26}$$

Next, from (8)

$$\mathbf{h} = \mathbf{R}^{-1}[\mathbf{r} - \mathbf{P}\mathbf{g}], \tag{27}$$

and hence,

$$\mathbf{h}^H\mathbf{s} = [\mathbf{r} - \mathbf{P}\mathbf{g}]^H\mathbf{R}^{-1}\mathbf{s} = \mathbf{r}^H\mathbf{R}^{-1}\mathbf{s} - \mathbf{g}^H\mathbf{P}^*\mathbf{R}^{-1}\mathbf{s}. \tag{28}$$

Upon substituting (28) and (12) into (26), the term $\triangle$CMSE can be further expressed as

$$\begin{aligned}\triangle\mathrm{CMSE} &= \mathbf{r}^H\mathbf{R}^{-1}\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^* - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{s} + \mathbf{g}^H[\mathbf{r}^* - \mathbf{P}^*\mathbf{R}^{-1}\mathbf{s}]\\ &= \mathbf{r}^H\mathbf{R}^{-1}[\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^* - \mathbf{s}] + [\mathbf{r}^H - \mathbf{s}^T\mathbf{R}^{-*}\mathbf{P}^*]\mathbf{g}^*\\ &= \{-\mathbf{r}^H\mathbf{R}^{-1} + [\mathbf{r}^H - \mathbf{s}^T\mathbf{R}^{-*}\mathbf{P}^*][\mathbf{R} - \mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1}\}\\ &\quad \cdot [\mathbf{s} - \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*].\end{aligned} \tag{29}$$

Remark 2: The $\triangle$CMSE in (29) is expressed in terms of the augmented complex statistics of the regressor $\mathbf{x}$ (both $\mathbf{R}$ and $\mathbf{P}$) and the joint augmented complex statistics of the estimandum y and the regressor $\mathbf{x}$ (both $\mathbf{r}$ and $\mathbf{s}$ in (10)). It serves as a complementary measure to the $\triangle$MSE in (21), to express a full second order performance advantage of the WL estimator.

IV. Advantages of a Joint Consideration of MSE Analysis and CMSE Analysis

In general, the complementary second order statistics offers enhanced degree of freedom to describe the second order statistics in real and imaginary data channels. To illustrate this advantage, first consider a scalar complex-valued random variable $x = x_r + \imath x_\imath$. Then, its variance and complementary variance are respectively given by

$$\sigma_x^2 = E[xx^*] = \sigma_r^2 + \sigma_\imath^2, \tag{30}$$

$$\widetilde{\sigma}_x^2 = E[x^2] = \sigma_r^2 - \sigma_\imath^2 + 2\imath\sigma_{r\imath}, \tag{31}$$

where $\sigma_r^2 = E[x_r^2]$, $\sigma_\imath^2 = E[x_\imath^2]$, and $\sigma_{r\imath} = E[x_r x_\imath]$. Observe that the complementary variance $\widetilde{\sigma}_x^2$ in (31) is complex-valued and has two degrees of freedom to describe the second order statistics in real and imaginary channels. On the other hand, the variance σ_x^2 in (30) contains information about the overall power in data, but it cannot tell us how the power is distributed across the real and imaginary channels due to its real-valued nature; a problem resulting from a single degree of freedom.

In a similar way, the proposed CMSE analysis enables us to quantify the degrees of improperness of the WL and SL estimation errors, e_{SL} and e_{WL}, as well as the difference in their degrees of improperness. Notice that the complex-valued nature of the complementary second order statistics provides the necessary two degrees of freedom to describe the improperness. This becomes clear when $\triangle$CMSE is expressed in terms of the real and imaginary parts of the respective estimation errors, to give

$$\begin{aligned}\triangle\text{CMSE} &= E[e_{\text{SL}}^2] - E[e_{\text{WL}}^2] \\ &= E[(e_{\text{SL},r} + \imath e_{\text{SL},\imath})^2] - E[(e_{\text{WL},r} + \imath e_{\text{WL},\imath})^2] \\ &= E[e_{\text{SL},r}^2] - E[e_{\text{SL},\imath}^2] + 2\imath E[e_{\text{SL},r} e_{\text{SL},\imath}] \\ &\quad - E[e_{\text{WL},r}^2] + E[e_{\text{WL},\imath}^2] - 2\imath E[e_{\text{WL},r} e_{\text{WL},\imath}] \\ &= \underbrace{E[e_{\text{SL},r}^2] - E[e_{\text{WL},r}^2]}_{\triangle\text{MSE}_r} - (\underbrace{E[e_{\text{SL},\imath}^2] - E[e_{\text{WL},\imath}^2]}_{\triangle\text{MSE}_\imath}) \\ &\quad + 2\imath(E[e_{\text{SL},r} e_{\text{SL},\imath}] - E[e_{\text{WL},r} e_{\text{WL},\imath}]). \end{aligned} \tag{32}$$

The real part, denoted by $\Re[\triangle\text{CMSE}]$, is of particular interest, since it contains information on the differences in MSE between the SL and WL estimators in the individual real and imaginary channels, denoted by $\triangle\text{MSE}_r$ and $\triangle\text{MSE}_\imath$, respectively; for more detail, see (34) and (35).

In a similar way, we can factorise $\triangle$MSE in (20) to yield

$$\begin{aligned}\triangle\text{MSE} &= E[|e_{\text{SL}}|^2] - E[|e_{\text{WL}}|^2] \\ &= E[e_{\text{SL},r}^2] + E[e_{\text{SL},\imath}^2] - E[e_{\text{WL},r}^2] - E[e_{\text{WL},\imath}^2] \\ &= \underbrace{E[e_{\text{SL},r}^2] - E[e_{\text{WL},r}^2]}_{\triangle\text{MSE}_r} + \underbrace{E[e_{\text{SL},\imath}^2] - E[e_{\text{WL},\imath}^2]}_{\triangle\text{MSE}_\imath}. \end{aligned} \tag{33}$$

Upon inspection of both sides of (32) and (33), we obtain

$$\triangle\text{MSE}_r = \frac{\triangle\text{MSE} + \Re[\triangle\text{CMSE}]}{2}, \tag{34}$$

$$\triangle\text{MSE}_\imath = \frac{\triangle\text{MSE} - \Re[\triangle\text{CMSE}]}{2}. \tag{35}$$

Remark 3: From (21), (29), (34) and (35), observe that only a joint consideration of the standard MSE and the proposed CMSE analyses provides enough degrees of freedom to move beyond just error power analysis, and to precisely quantify the MSE performance differences between the SL and WL estimators in both the real and imaginary channels, that is, $\triangle\text{MSE}_r$ and $\triangle\text{MSE}_\imath$. In general, the standard MSE analysis [8] reflects only error power difference, and is not sufficient to model its distribution across data channels.

V. Joint Consideration of MSE Analysis and CMSE Analysis by Case Studies

After establishing the advantages of CMSE analysis, it is natural to ask whether a WL estimator offers individual MSE performance advantages in both the real and imaginary channels over its SL counterpart. The analysis in Section IV indicates that this can be achieved by investigating the relationship between $\triangle$MSE and its associated CMSE difference, $\triangle$CMSE, of both WL and SL estimators. For rigour, we next consider three case studies, covering all the characteristic situations in the complex domain $\mathbb{C}$.

A. Case #1: Strictly Linear Estimation, Where $\mathbf{s} = \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$

From (21) and (29), this condition immediately results in

$$\triangle\text{MSE} = \triangle\text{CMSE} = 0, \tag{36}$$

and consequently, from (34) and (35), we have

$$\triangle\text{MSE}_r = \triangle\text{MSE}_\imath = 0. \tag{37}$$

This is as expected, because the optimal weight vector $\mathbf{g}$ in (12), obtained by the WL estimator, vanishes, subject to this condition, and after a few mathematical manipulations, it can be shown that $\mathbf{h}$ in (11) is equal to $\mathbf{h} = \mathbf{R}^{-1}\mathbf{r}$, which is consistent with (17), obtained by the SL estimator. Therefore, the WL estimator in (2) reduces to the SL one in (1), with identical MSE performances in both the real and imaginary channels and the same degree of improperness in the estimation error. A special case of this condition is the so called *jointly circular case* [8], characterised by $\mathbf{s} = \mathbf{0}$ and $\mathbf{P} = \mathbf{0}$, which is precisely the statistical convenience which is explicitly or implicitly adopted in traditional estimation problems in the complex domain $\mathbb{C}$.

B. Case #2: A Jointly Improper Pair $\{y, \mathbf{x}\}$ with a Proper Regressor $\mathbf{x}$, where $\mathbf{s} \neq \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ and $\mathbf{P} = \mathbf{0}$

The above conditions can be simplified to $\mathbf{s} \neq \mathbf{0}$ and $\mathbf{P} = \mathbf{0}$, indicating that the properness is only valid for the regressor $\mathbf{x}$, characterised by $\mathbf{P} = \mathbf{0}$, and with no assumption imposed on the estimandum y. The expressions for the optimal weight vectors $\mathbf{h}$ and $\mathbf{g}$ in (11) and (12) can now be respectively simplified into

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{r}, \;\; \mathbf{g} = \mathbf{R}^{-*}\mathbf{s}^*. \tag{38}$$

Note that the term $\mathbf{h} = \mathbf{R}^{-1}\mathbf{r}$ remains the same as that in (17), obtained by using the SL estimator. This stems from the properness assumption on the regressor $\mathbf{x}$, which implies that $\mathbf{x}$ and $\mathbf{x}^*$ are uncorrelated, i.e., $E[\mathbf{x}(\mathbf{x}^*)^H] = E[\mathbf{x}\mathbf{x}^T] = \mathbf{P} = \mathbf{0}$. Therefore, within the WL estimator in (2), the Hilbert subspaces generated by $\mathbf{x}$ and $\mathbf{x}^*$, expressed in (4)–(7), are orthogonal, since $\mathbf{x}^*$ does not affect the term arising from $\mathbf{x}$, and *vice versa*. This leads to the vanishing of the vector $\mathbf{a}$ in (29), and explains the simplifications of (21) and (29), given by

$$\triangle\text{MSE} = \mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}, \; \triangle\text{CMSE} = 0. \tag{39}$$

Consequently, from (34) and (35), we obtain

$$\triangle\text{MSE}_r = \triangle\text{MSE}_\imath = \frac{\triangle\text{MSE}}{2} = \frac{1}{2}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}. \tag{40}$$

Therefore, a nonzero vector $\mathbf{s}$ necessarily implies that there still exists an MSE performance advantage when using the WL estimator instead of the SL one, quantified by $\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}$, and **the corresponding identical degrees of improperness in both SL and WL estimators further indicate that this advantage is equally carried in the real and imaginary channels**.

C. Case #3: A Jointly Improper Pair $\{y, \mathbf{x}\}$ with an Improper Regressor $\mathbf{x}$, where $\mathbf{s} \neq \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ and $\mathbf{P} \neq \mathbf{0}$

For this most general case, in order to develop the relation between $\triangle$MSE and $\triangle$CMSE of both the SL and WL estimators, we consider a more explicit expression of $\triangle$CMSE in (29) by first evaluating the vector within the $\{\cdot\}$ operator as

$$\begin{aligned}-\mathbf{r}^H\mathbf{R}^{-1} &+ [\mathbf{r}^H - \mathbf{s}^T\mathbf{R}^{-*}\mathbf{P}^*][\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1} \\ &= -\mathbf{r}^H\mathbf{R}^{-1}[\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*][\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1} \\ &\quad + [\mathbf{r}^H - \mathbf{s}^T\mathbf{R}^{-*}\mathbf{P}^*][\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1} \\ &= -[\mathbf{s}^T - \mathbf{r}^H\mathbf{R}^{-1}\mathbf{P}]\mathbf{R}^{-*}\mathbf{P}^*[\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1} \\ &= -[\mathbf{s}-\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*]^T\mathbf{R}^{-*}\mathbf{P}^*[\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1}.\end{aligned} \tag{41}$$

Now, a substitution of (41) into (29) yields

$$\begin{aligned}&\triangle\text{CMSE} \\ &= -[\mathbf{s}-\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*]^T\mathbf{R}^{-*}\mathbf{P}^*[\mathbf{R}-\mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*]^{-1}[\mathbf{s}-\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*].\end{aligned} \tag{42}$$

This expression is now much more similar to that for $\triangle$MSE in (21), in the sense that both are a quadratic form of the non-zero vector $\mathbf{s}-\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$.

Theorem 1: The $|\triangle\text{CMSE}| < \triangle\text{MSE}$, subject to the conditions $\mathbf{s} \neq \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ and $\mathbf{P} \neq \mathbf{0}$.

Proof: First, consider a joint diagonalisation of $\mathbf{R}$ and $\mathbf{P}$ by using the strong uncorrelating transform (SUT) [39]–[41], given by

$$\mathbf{R} = \mathbf{C}\mathbf{I}\mathbf{C}^H, \ \mathbf{P} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^T, \tag{43}$$

where $\mathbf{C}$ is an $N \times N$ nonsingular square matrix, the inverse of which, $\mathbf{C}^{-1}$, exists and is known as the SUT matrix, $\mathbf{I}$ is the identity matrix and $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ is a diagonal matrix of the circularity coefficients, whose values satisfy $1 \geq \lambda_1 \geq \cdots \geq \lambda_N \geq 0$.

By using SUT in (43), the non-zero vector $\mathbf{s}-\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ in both (21) and (42) can be expressed as

$$\mathbf{s}-\mathbf{P}\mathbf{R}^{-*}\mathbf{r}^* = \mathbf{s}-\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*, \tag{44}$$

and according to the analysis provided in Appendix A, upon such joint diagonalisation, the $\triangle$MSE in (21) and $\triangle$CMSE in (42) can be expressed as

$$\triangle\text{MSE} = [\mathbf{C}^{-1}\mathbf{s}-\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*]^H\frac{\mathbf{I}}{\mathbf{I}-\boldsymbol{\Lambda}^2}[\mathbf{C}^{-1}\mathbf{s}-\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*], \tag{45}$$

$$\triangle\text{CMSE} = [\mathbf{C}^{-1}\mathbf{s}-\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*]^T\frac{-\boldsymbol{\Lambda}}{\mathbf{I}-\boldsymbol{\Lambda}^2}[\mathbf{C}^{-1}\mathbf{s}-\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*]. \tag{46}$$

Remark 4: Note that in (43), the condition $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 1$ holds only when $\mathbf{x}$ is real-valued, i.e., maximum noncircular with $\mathbf{R} = \mathbf{P}$ [39], [42]. In other words, for any complex-valued random vector $\mathbf{x}$ discussed above, the matrix $\mathbf{I}-\boldsymbol{\Lambda}^2$ above is always nonsingular.

Now, by defining a non-zero $N \times 1$ vector $\mathbf{e}$ as

$$\mathbf{e} = \mathbf{C}^{-1}\mathbf{s}-\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*, \tag{47}$$

where $\mathbf{e} = [e_1, e_2, \ldots, e_N]^T$, and by considering (45) and (46), we have

$$\begin{aligned}|\triangle\text{CMSE}| &= \left|\mathbf{e}^T\frac{\boldsymbol{\Lambda}}{\mathbf{I}-\boldsymbol{\Lambda}^2}\mathbf{e}\right| = \sum_{n=1}^{N}\frac{\lambda_n\left|e_n^2\right|}{1-\lambda_n^2} \\ &< \sum_{n=1}^{N}\frac{\left|e_n^2\right|}{1-\lambda_n^2} \leq \sum_{n=1}^{N}\frac{\left|e_n\right|^2}{1-\lambda_n^2} \\ &= \mathbf{e}^H\frac{\mathbf{I}}{\mathbf{I}-\boldsymbol{\Lambda}^2}\mathbf{e} = \triangle\text{MSE}.\end{aligned} \tag{48}$$

This completes the proof of Theorem 1. ■

Corollary 1: Subject to the conditions $\mathbf{s} \neq \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ and $\mathbf{P} \neq \mathbf{0}$, the bound $|\Re[\triangle\text{CMSE}]| < \triangle\text{MSE}$ holds.

Proof: The proof follows directly from Theorem 1, since $|\Re[\triangle\text{CMSE}]| \leq |\triangle\text{CMSE}|$. ■

Corollary 2: Subject to the conditions $\mathbf{s} \neq \mathbf{P}\mathbf{R}^{-*}\mathbf{r}^*$ and $\mathbf{P} \neq \mathbf{0}$, it follows that $\triangle\text{MSE}_r > 0$ and $\triangle\text{MSE}_\imath > 0$.

Proof: The proof follows directly from Corollary 1, and (34) and (35). ■

D. Numerical Evaluation

Numerical experiments in a widely linear system identification setting, as described in (2), were conducted in the MATLAB programming environment to evaluate theoretical findings in the above case studies. The widely linear system coefficients to be identified were those typically used to describe frequency-dependent inphase/quadrature (I/Q) imbalance in the direct-conversion transceivers for wideband wireless systems, given by [24], [25]

$$\mathbf{h}_\text{o} = \frac{\mathbf{a}+\gamma e^{-\jmath\theta}\mathbf{b}}{2}, \ \mathbf{g}_\text{o} = \frac{\mathbf{a}-\gamma e^{\jmath\theta}\mathbf{b}}{2}, \tag{49}$$

where $\mathbf{a} = [0.01, 1, 0.01]^T$ and $\mathbf{b} = [0.01, 1, 0.2]^T$ are respectively the low-pass filter coefficients for the I (real) and Q (imaginary) branches of the transceiver, and the gain mismatch and phase mismatch between the two branches were respectively $\gamma = 1.05$ and $\theta = 8^\circ$. The desired signal (estimandum) y at time index k, that is, $y(k)$, was generated as

$$y(k) = \mathbf{h}_\text{o}^H\mathbf{x}(k) + \mathbf{g}_\text{o}^H\mathbf{x}^*(k) + q(k), \tag{50}$$

where the system input $\mathbf{x}(k)$ was initially an improper Gaussian autoregressive AR(1) process, given by

$$x(k) = 0.9x(k-1) + n(k), \tag{51}$$

while the driving noise $n(k)$ was a zero-mean doubly white improper Gaussian process with variance $\sigma_n^2 = 1$ and complementary variance $\tilde{\sigma}_n^2 = 0.9$. The system noise $q(k)$ in (50) was a zero-mean doubly white proper Gaussian process, whose variance σ_q^2 was chosen so as to give the signal-to-noise ratio at 20 dB. All the simulation results were obtained by averaging over 10,000 independent trials.

This particular system setting satisfied the conditions of Case #3 in Section V-C in the sense that $\{y(k), \mathbf{x}(k)\}$ were a jointly improper pair with an improper regressor $\mathbf{x}(k)$. The theoretical MSE and complementary MSE (CMSE) differences between

TABLE I
THEORETICAL AND SIMULATED MSE AND CMSE DIFFERENCES BETWEEN THE SL AND WL ESTIMATORS IN DIFFERENT CASE STUDIES

	$\triangle$MSE		$\triangle$CMSE		$\triangle\text{MSE}_r$		$\triangle\text{MSE}_\imath$	
	theoretical (21)	simulated	theoretical (29)	simulated	theoretical (34)	simulated	theoretical (35)	simulated
Case #3	0.0169	0.0170	-0.0044 + $\imath$0.0137	-0.0046 + $\imath$0.0138	0.0062	0.0062	0.0106	0.0108
Case #2	0.0894	0.0886	0	0	0.0447	0.0443	0.0447	0.0443
Case #1	0	0	0	0	0	0	0	0

the SL and WL estimators, that is, $\triangle$MSE and $\triangle$CMSE, were respectively evaluated by using (21) and (29) and are given in Table I, while their MSE differences in the real and imaginary channels, that is, $\triangle\text{MSE}_r$ and $\triangle\text{MSE}_\imath$, were subsequently obtained by using (34) and (35), respectively. It has been justified by both the theoretical and simulated evaluations that the inequality $|\triangle\text{CMSE}| < \triangle\text{MSE}$ does hold, as proved in Theorem 1. Consequently, according to Corollary 2, we can draw the conclusion that the WL estimator has performance advantages in terms of MSE over its SL counterpart in both I (real) and Q (imaginary) channels, since $\triangle\text{MSE}_r > 0$ and $\triangle\text{MSE}_\imath > 0$, as evidenced by both the theoretical and simulated results in Table I.

We next considered a more constrained Case #2 in Section V-B, whose conditions were satisfied by making the Gaussian input $\mathbf{x}(k)$ to the widely linear system in (50) proper. In the experiment, this was achieved by using a proper driving noise $q(k)$ with a vanishing complementary variance $\tilde{\sigma}_n^2$ in (51). As discussed in Section V-B, this further constraint led to the degrees of the improperness of both estimators being identical, as indicated by $\triangle\text{CMSE} = 0$, and hence, the I and Q channels carried a half of the total MSE advantages of the WL estimator over the SL one, evidenced by $\triangle\text{MSE}_r = \triangle\text{MSE}_\imath = \triangle\text{MSE}/2$.

In the final stage, we investigated the most stringent conditions, as stated in Case #1 in Section V-A. These conditions were achieved by further setting $\mathbf{b} = \mathbf{a} = [0.01, 1, 0.01]^T$, $\gamma = 1$ and $\theta = 0^\circ$ alongside the proper regressor $\mathbf{x}(k)$. In this way, the WL system identification task in (50) reduced to an SL one. As expected, in this situation the WL estimator could not offer any performance advantage as compared with the SL estimator, evidenced by $\triangle\text{MSE} = \triangle\text{CMSE} = \triangle\text{MSE}_r = \triangle\text{MSE}_\imath = 0$. Table I shows that in all the case studies considered above, the simulated results closely matched their theoretical counterparts.

VI. DISCUSSION AND CONCLUSION

Through a joint analysis of the standard MSE and the proposed complementary MSE (CMSE), we have provided new important and useful interpretations of full second order behaviours of the SL and WL estimators, which have been missing in the open literature. The proposed complementary MSE (CMSE) analysis has illustrated that the CMSE (or improperness) difference, characterised by $\triangle$CMSE, between the SL and WL estimators, is bounded by its associated MSE difference, in the sense that $|\triangle\text{CMSE}| \leq \triangle\text{MSE}$. The equality holds only when $\mathbf{s} = \mathbf{PR}^{-*}\mathbf{r}^*$, a subcategory of which is the so-called *jointly circular case*, characterised by $\mathbf{s} = \mathbf{0}$ and $\mathbf{P} = \mathbf{0}$. Secondly, a joint consideration of the standard MSE analysis and the proposed CMSE analysis has illustrated that the WL estimator always yields performance advantages over its SL counterpart in both the real and imaginary channels, since $\triangle\text{MSE}_r > 0$ and $\triangle\text{MSE}_\imath > 0$, except for the special condition $\mathbf{s} = \mathbf{PR}^{-*}\mathbf{r}^*$ for which $\triangle\text{MSE}_r = \triangle\text{MSE}_\imath = 0$. Note that this rigourous characterisation cannot be achieved by using the standard MSE analysis only, due to its insufficient degrees of freedom to describe both the full second order statistics of data and the parameters of both SL and WL estimators. Thirdly, since the concept of improperness is a second order statistical property of random vectors in division algebras beyond the real domain $\mathbb{R}$, the physical intuition behind WL and SL estimators, as illustrated by the proposed analysis, is not limited to the complex domain $\mathbb{C}$ and can be generalised to other hyper-complex domains, e.g., the quaternion domain $\mathbb{H}$ and the octonion domain $\mathbb{O}$. For example, the overall MSE performance advantage, along all the four dimensions in $\mathbb{H}$, of the quaternion WL estimator over its SL counterpart, has been currently rigourously quantified for the generality of quaternion-valued Gaussian data, only in the standard way [43]–[46]. Simulations in the system identification setting support the analysis.

APPENDIX A
DETAILED DERIVATION OF $\triangle$MSE IN (21) AND $\triangle$CMSE IN (42) BY USING SUT

Consider the covariance matrix of the augmented input vector $\mathbf{x}^a = [\mathbf{x}^T, \mathbf{x}^H]^T$, defined as [4], [6], [7]

$$\begin{aligned}\mathbf{R}^a &= E[\mathbf{x}^a\mathbf{x}^{aH}] \\ &= \begin{bmatrix} E[\mathbf{x}\mathbf{x}^H] & E[\mathbf{x}\mathbf{x}^T] \\ E[\mathbf{x}^*\mathbf{x}^H] & E[\mathbf{x}^*\mathbf{x}^T] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{R}^* \end{bmatrix}. \end{aligned} \tag{52}$$

Note that its inverse $\mathbf{R}^{-a}$ can be expressed as [20]

$$\mathbf{R}^{-a} = \begin{bmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{D}^* & \mathbf{A}^* \end{bmatrix}, \tag{53}$$

where the $N \times N$ Hermitian matrix $\mathbf{A}$ and the symmetric matrix $\mathbf{D}$ are given by [20]

$$\mathbf{A} = [\mathbf{R} - \mathbf{PR}^{-*}\mathbf{P}^*]^{-1}, \ \mathbf{D} = -\mathbf{APR}^{-*}. \tag{54}$$

Now, by using the Hermitian nature of $\mathbf{A}$ and $\mathbf{R}$, as well as the symmetric nature of $\mathbf{D}$ and $\mathbf{P}$, from (54), we further have

$$\mathbf{D} = \mathbf{D}^T = -\mathbf{R}^{-1}\mathbf{PA}^* = -\mathbf{R}^{-1}\mathbf{P}[\mathbf{R} - \mathbf{PR}^{-*}\mathbf{P}^*]^{-*}. \tag{55}$$

Therefore, based on (44), (54) and (55), from (21) and (42), we have

$$\triangle\text{MSE} = [\mathbf{s} - \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*]^H \mathbf{A}[\mathbf{s} - \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*], \quad (56)$$

$$\triangle\text{CMSE} = [\mathbf{s} - \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*]^T \mathbf{D}^*[\mathbf{s} - \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{-*}\mathbf{r}^*]. \quad (57)$$

We can now use the SUT diagonalised covariance and complementary matrices in (43) to represent $\mathbf{R}^a$ in (52), so that

$$\begin{aligned}\mathbf{R}^a &= \begin{bmatrix} \mathbf{CIC^H} & \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^T \\ \mathbf{C}^*\boldsymbol{\Lambda}\mathbf{C}^H & \mathbf{C}^*\mathbf{IC}^T \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^* \end{bmatrix}}_{\mathbf{Q}} \begin{bmatrix} \mathbf{I} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \mathbf{I} \end{bmatrix} \underbrace{\begin{bmatrix} \mathbf{C^H} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^T \end{bmatrix}}_{\mathbf{Q}^H}. \end{aligned} \quad (58)$$

In the same spirit, after a few mathematical manipulations, we have

$$\begin{aligned}&\mathbf{R}^{-a} \\ &= \underbrace{\begin{bmatrix} \mathbf{C^{-H}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-T} \end{bmatrix}}_{\mathbf{Q}^{-H}} \frac{\begin{bmatrix} \mathbf{I} & -\boldsymbol{\Lambda} \\ -\boldsymbol{\Lambda} & \mathbf{I} \end{bmatrix}}{\mathbf{I} - \boldsymbol{\Lambda}^2} \underbrace{\begin{bmatrix} \mathbf{C^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{-*} \end{bmatrix}}_{\mathbf{Q}^{-1}} \\ &= \frac{\begin{bmatrix} \mathbf{C^{-H}IC^{-1}} & -\mathbf{C}^{-H}\boldsymbol{\Lambda}\mathbf{C}^{-*} \\ -\mathbf{C}^{-T}\boldsymbol{\Lambda}\mathbf{C}^{-1} & \mathbf{C}^{-T}\mathbf{IC}^{-*} \end{bmatrix}}{\mathbf{I} - \boldsymbol{\Lambda}^2}, \end{aligned} \quad (59)$$

where $\mathbf{C}^{-H}$ and $\mathbf{C}^{-T}$ respective denote $(\mathbf{C}^{-1})^H$ and $(\mathbf{C}^{-1})^T$. By comparing (53) and (59), we obtain

$$\mathbf{A} = \mathbf{C}^{-H}\frac{\mathbf{I}}{\mathbf{I} - \boldsymbol{\Lambda}^2}\mathbf{C}^{-1}, \ \mathbf{D}^* = \mathbf{C}^{-T}\frac{-\boldsymbol{\Lambda}}{\mathbf{I} - \boldsymbol{\Lambda}^2}\mathbf{C}^{-1}, \quad (60)$$

and upon substituting these back into (56) and (57), this yields (45) and (46).

References

[1] F. D. Neeser and J. L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[2] B. Picinbono, “On circularity,” *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York, NY, USA: Wiley, 2003.

[4] B. Picinbono, “Second-order complex random vectors and normal distributions,” *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.

[5] P. J. Schreier and L. L. Scharf, “Second-order analysis of improper complex random vectors and process,” *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[6] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York, NY, USA: Wiley, 2009.

[7] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[8] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[9] S. C. Douglas, “Widely-linear recursive least-squares algorithm for adaptive beamforming,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2009, pp. 2041–2044.

[10] J. Navarro-Moreno, J. Moreno-Kayser, R. Fernandez-Alcala, and J. Ruiz-Molina, “Widely linear estimation algorithms for second-order stationary signals,” *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4930–4935, Dec. 2009.

[11] J. Navarro-Moreno, M. Estudillo-Martinez, R. Fernandez-Alcala, and J. Ruiz-Molina, “Estimation of improper complex-valued random signals in colored noise by using the Hilbert space theory,” *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2859–2867, Jun. 2009.

[12] S. C. Douglas and D. P. Mandic, “Performance analysis of the conventional complex LMS and augmented complex LMS algorithms,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2010, pp. 3794–3797.

[13] Y. Xia, D. P. Mandic, and A. H. Sayed, “An adaptive diffusion augmented CLMS algorithm for distributed filtering of noncircular complex signals,” *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 659–662, Nov. 2011.

[14] B. Jelfs, D. P. Mandic, and S. C. Douglas, “An adaptive approach for the identification of improper complex signals,” *Signal Process.*, vol. 92, no. 2, pp. 335–344, 2012.

[15] D. Dini and D. P. Mandic, “Class of widely linear complex Kalman filters,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 775–786, May 2012.

[16] D. P. Mandic, S. Kanna, and S. C. Douglas, “Mean square analysis of the CLMS and ACLMS for non-circular signals: The approximate uncorrelating transform approach,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 3531–3535.

[17] Y. Xia, S. C. Douglas, and D. P. Mandic, “Performance analysis of the deficient length augmented CLMS algorithm for second order noncircular complex signals,” *Signal Process.*, vol. 144, pp. 214–225, Mar. 2018.

[18] W. H. Gerstacker, R. Schober, and A. Lampe, “Receivers with widely linear processing for frequency-selective channels,” *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1512–1523, Nov. 2003.

[19] P. Chevalier and F. Pipon, “New insights into optimal widely linear array receivers for the demodulation of BPSK, MSK and GMSK signals corrupted by noncircular interferences–Application to SAIC,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 870–883, Mar. 2006.

[20] P. Chevalier and A. Blin, “Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences,” *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5323–5336, Nov. 2007.

[21] T. Fusco and M. Tanda, “Blind frequency-offset estimation for OFDM/OQAM systems,” *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1828–1838, May 2007.

[22] Y.-M. Shi, L. Huang, C. Qian, and H. C. So, “Shrinkage linear and widely linear complex-valued least mean squares algorithms for adaptive beamforming,” *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 119–131, Jan. 2015.

[23] M. Valkama, M. Renfors, and V. Koivunen, “Blind signal estimation in conjugate signal models with application to I/Q imbalance compensation,” *IEEE Signal Process. Lett.*, vol. 12, no. 11, pp. 733–736, Nov. 2005.

[24] L. Anttila, M. Valkama, and M. Renfors, “Circularity-based I/Q imbalance compensation in wideband direct-conversion receivers,” *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2099–2113, Jul. 2008.

[25] Y. Tsai, C.-P. Yen, and X. Wang, “Blind frequency-dependent I/Q imbalance compensation for direct-conversion receivers,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 1976–1986, Jun. 2010.

[26] A. Hakkarainen, J. Werner, K. R. Dandekar, and M. Valkama, “Widely-linear beamforming and RF impairment suppression in massive antenna arrays,” *J. Commun. Netw.*, vol. 15, no. 4, pp. 383–397, Aug. 2013.

[27] A. Gokceoglu, S. Dikmese, M. Valkama, and M. Renfors, “Energy detection under IQ imbalance with single- and multi-channel direct-conversion receiver: Analysis and mitigation,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 3, pp. 411–424, Mar. 2014.

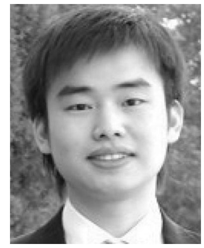
[28] Z. Li, Y. Xia, W. Pei, K. Wang, Y. Huang, and D. P. Mandic, “Noncircular measurement and mitigation of I/Q imbalance for OFDM-based WLAN transmitters,” *IEEE Trans. Instrum. Meas.*, vol. 66, no. 3, pp. 383–393, Mar. 2017.

[29] Z. Li, Y. Xia, W. Pei, K. Wang, and D. P. Mandic, “An augmented nonlinear complex LMS for digital self-interference cancellation in full-duplex direct-conversion transceivers,” *IEEE Trans. Signal Process.*, submitted for publication.

[30] Y. Xia, S. C. Douglas, and D. P. Mandic, “Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators,” *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 44–54, Sep. 2012.

[31] Y. Xia and D. P. Mandic, “Widely linear adaptive frequency estimation of unbalanced three-phase power systems,” *IEEE Trans. Instrum. Meas.*, vol. 61, no. 1, pp. 74–83, Jan. 2012.

[32] D. H. Dini and D. P. Mandic, "Widely linear modeling for frequency estimation in unbalanced three-phase power systems," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 2, pp. 353–363, Feb. 2013.
[33] S. Javidi, D. P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second-order statistics," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 7, pp. 1404–1416, Oct. 2010.
[34] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex statistics," *Renewable Energy*, vol. 34, no. 1, pp. 44–54, Jan. 2009.
[35] Y. Xia, B. Jelfs, M. M. Van Hulle, J. C. Principe, and D. P. Mandic, "An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 74–83, Jan. 2011.
[36] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008.
[37] Y. Xia and D. P. Mandic, "Complementary mean square analysis of augmented CLMS for second-order noncircular Gaussian signals," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1413–1417, Sep. 2017.
[38] Y. Xia and D. P. Mandic, "A full mean square analysis of CLMS for second-order noncircular inputs," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5578–5590, Nov. 2017.
[39] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1017–1029, Mar. 2006.
[40] S. C. Douglas, "Fixed-point algorithms for the blind separation of arbitrary complex-valued non-Gaussian signal mixtures," *EURASIP J. Adv. Signal Process.*, vol. 2007, pp. 1–15, 2007.
[41] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Process.*, vol. 89, pp. 365–377, 2009.
[42] C. Cheong Took, S. C. Douglas, and D. P. Mandic, "On approximate diagonalization of correlation matrices in widely linear signal processing," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1469–1473, Mar. 2012.
[43] C. Cheong Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4427–4431, Aug. 2010.
[44] J. Navarro-Moreno, R. M. Fernandez-Alcala, and J. C. Ruiz-Molina, "A quaternion widely linear model for nonlinear Gaussian estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6414–6424, Dec. 2014.
[45] Y. Xia, C. Jahanchahi, T. Nitta, and D. P. Mandic, "Peformance bounds of quaternion estimators," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3287–3292, Dec. 2015.
[46] X. Min, S. Kanna, S. C. Douglas, and D. P. Mandic, "Performance advantage of quaternion widely linear estimation: An approximate uncorrelating transform approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 4348–4352.



Yili Xia (M'11) received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2006, and the M.Sc. (Distinction) degree in communications & signal processing from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2007. He received the Ph.D. degree in adaptive signal processing from the same college in 2011.

Since 2013, he has been an Associate Professor with the School of Information and Engineering, Southeast University, Nanjing, China, where he is currently the Deputy Head of the Department of Information and Signal Processing Engineering. His research interests include complex and hyper-complex statistical analysis, linear and nonlinear adaptive filters, as well as their applications on communications and power systems.

Danilo P. Mandic (M'99–SM'03–F'12) received the Ph.D. degree in nonlinear adaptive signal processing from Imperial College London, London, U.K., in 1999, where he is currently a Professor in signal processing.

He has been working in the area of nonlinear adaptive signal processing, multivariate data analysis, and nonlinear dynamics. He has been a Guest Professor at Katholieke Universiteit Leuven, Leuven, Belgium, the Tokyo University of Agriculture and Technology, Tokyo, Japan, and Westminster University, London, and a Frontier Researcher at RIKEN, Japan. His publication record includes two research monographs entitled *Recurrent Neural Networks for Prediction*: *Learning Algorithms*, *Architectures and Stability* (1st ed., Aug. 2001) and *Complex Valued Nonlinear Adaptive Filters*: *Noncircularity*, *Widely Linear and Neural Models* (1st ed., Wiley, Apr. 2009), an edited book titled *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (Springer, 2008) and more than 200 publications on signal and image processing.

Prof. Mandic has been a member of the IEEE Technical Committee on Signal Processing Theory and Methods, and has been an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS, and the *International Journal of Mathematical Modelling and Algorithms*. He is currently an Associate Editor for the IEEE SIGNAL PROCESSING MAGAZINE and the IEEE TRANSACTIONS ON INFORMATION AND SIGNAL PROCESSING OVER NETWORKS. He has produced award winning papers and products resulting from his collaboration with Industry.