Component-Wise Conditionally Unbiased Widely Linear MMSE Estimation

Mario Huemer, *Senior Member, IEEE,* Oliver Lang, *Student Member, IEEE,*
Christian Hofbauer, *Student Member, IEEE*

***Abstract*—Biased estimators can outperform unbiased ones in terms of the mean square error (MSE). In this work we treat all estimators in the Bayesian framework, where the best linear unbiased estimator (BLUE) fulfills the so called global conditional unbiased constraint. Recently, component-wise conditionally unbiased linear minimum mean square error (CWCU LMMSE) estimators have been introduced. These estimators preserve a quite strong (namely the CWCU) unbiasedness condition which in effect sufficiently represents the intuitive view of unbiasedness, while in fact they are in general global conditionally biased. Overall, CWCU LMMSE estimators constitute an interesting compromise between the BLUE and the LMMSE estimator. We briefly recapitulate CWCU LMMSE estimation under linear model assumptions, and additionally derive the CWCU LMMSE estimator under the (only) assumption of jointly Gaussian parameters and measurements. The main intent of this work, however, is the extension of the theory of CWCU estimation to CWCU widely linear estimators. We derive the CWCU WLMMSE estimator for different model assumptions and address the analytical relationships between the CWCU WLMMSE and the WLMMSE estimators. The properties of CWCU WLMMSE estimators are deduced analytically, and compared to global conditionally unbiased as well as WLMMSE counterparts with the help of a parameter estimation application.**

***Index Terms*—Bayesian Estimation, Best Linear Unbiased Estimator, Linear Minimum Mean Square Error, Widely Linear Estimation, component-wise conditionally unbiased, CWCU**

I. Introduction

USUALLY, when we talk about unbiased estimation of a parameter vector $\mathbf{x} \in \mathbb{C}^n$ out of a measurement vector $\mathbf{y} \in \mathbb{C}^m$, then the estimation problem is treated in the classical framework, where $\mathbf{x}$ is treated as deterministic but unknown [1]- [4]. Letting $\hat{\mathbf{x}} = \mathbf{g}(\mathbf{y})$ be an estimator of $\mathbf{x}$, then the classical unbiased constraint asserts that

$$E_{\mathbf{y}}[\hat{\mathbf{x}}] = \int \mathbf{g}(\mathbf{y}) p(\mathbf{y};\mathbf{x}) d\mathbf{y} = \mathbf{x} \quad \text{for all possible } \mathbf{x}, \tag{1}$$

where $p(\mathbf{y};\mathbf{x})$ is the probability density function (PDF) of vector $\mathbf{y}$ parametrized by the unknown parameter vector $\mathbf{x}$. The index of the expectation operator shall indicate the PDF over which the averaging is performed. In the Bayesian approach on the other hand $\mathbf{x}$ is treated as a random vector. The Bayesian unbiased constraint is

$$E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}} - \mathbf{x}] = \iint (\mathbf{g}(\mathbf{y}) - \mathbf{x})\, p(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y} = \mathbf{0}, \tag{2}$$

M. Huemer and O. Lang are with the Institute of Signal Processing, Johannes Kepler University, Linz, Austria. C. Hofbauer is with Linz Center of Mechatronics GmbH, Linz. E-mail: mario.huemer@jku.at, oliver.lang@jku.at, christian.hofbauer@lcm.at

This work has been supported by the Austrian Science Fund (FWF): I683-N13.

where the integration is performed over the joint PDF of $\mathbf{x}$ and $\mathbf{y}$. Compared to the classical unbiased constraint in (1), the Bayesian unbiased constraint is a much softer requirement, which will be particularly discussed in Section VII. However, Bayesian estimators in general allow to incorporate prior knowledge about the statistics of $\mathbf{x}$.

Eq. (1) can also be formulated in the Bayesian framework. Here, the corresponding problem arises by demanding global conditional unbiasedness, i.e.

$$E_{\mathbf{y}|\mathbf{x}}[\hat{\mathbf{x}}|\mathbf{x}] = \int \mathbf{g}(\mathbf{y}) p(\mathbf{y}|\mathbf{x}) d\mathbf{y} = \mathbf{x} \quad \text{for all possible } \mathbf{x}. \tag{3}$$

The attribute *global* indicates that the condition is made on the whole parameter vector $\mathbf{x}$. However, the constricting requirement in (3) prevents the exploitation of prior knowledge about the parameters, and hence leads to a significant reduction in the benefits brought about by the Bayesian framework.

In component-wise conditionally unbiased (CWCU) Bayesian parameter estimation [5]- [9], instead of constraining the estimator to be globally unbiased, we aim for achieving conditional unbiasedness on one parameter component at a time. Let x_i be the i^{th} element of $\mathbf{x}$, and $\hat{x}_i = g_i(\mathbf{y})$ be an estimator of x_i. Then the CWCU constraints are

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \int g_i(\mathbf{y}) p(\mathbf{y}|x_i) d\mathbf{y} = x_i, \tag{4}$$

for all possible x_i (and all $i = 1, 2, ..., n$). The CWCU constraints are less stringent than the global conditional unbiasedness condition in (3), and it will turn out that a CWCU estimator in many cases allows the incorporation of prior knowledge about the statistical properties of the parameter vector. In the following we denote the linear estimator fulfilling the CWCU constraints and minimizing the Bayesian mean square error (BMSE) the CWCU linear minimum mean square error (CWCU LMMSE) estimator. The CWCU LMMSE estimator cannot outperform the LMMSE estimator in a BMSE sense since it minimizes the BMSE under the additional constraints in (4), while the LMMSE estimator's only restriction is the linearity constraint. However, the CWCU estimators feature their inherent conditional unbiasedness property, which is visualized for a particular example in Fig. 1 (taken from [9]). In this example channel distorted and noisy received quadrature amplitude modulated (QAM) data symbols are estimated by the best linear unbiased estimator (BLUE), which fulfills (1), (2) and (4), the CWCU LMMSE estimator which fulfills (2) and (4), and the LMMSE estimator which only fulfills the weakest constraint (2). Fig. 1 shows the relative frequencies of the corresponding estimates in the complex

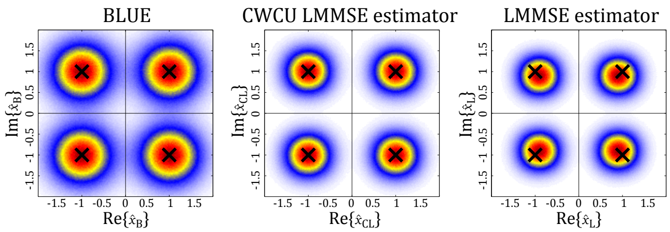

Fig. 1. Visualization of the relative frequencies of the BLUE, the CWCU LMMSE estimator, and the LMMSE estimator, respectively. The black crosses mark the ideal 4-QAM constellation points.

plane. The BLUE and the CWCU LMMSE estimator have their estimates centered around the true constellation points since these estimators fulfill the CWCU constraints. Note that in Fig. 1 the BMSE of the CWCU LMMSE estimator is clearly below the one of the BLUE. This further advantage is due to the fact that the CWCU constraints allow for the incorporation of prior knowledge about the data in this example. The LMMSE estimator is conditionally biased towards the prior mean which is 0. For details on that example we refer the reader to [9]. More examples and beneficial applications of the CWCU LMMSE estimator can be found in [5]- [7].

The theory of the CWCU LMMSE estimator under linear model assumptions has been discussed in [8]- [9]. The estimator is of the form $\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{b}$ with appropriate sized matrix $\mathbf{E}$ and vector $\mathbf{b}$, and it is mainly designed for proper measurement vectors. For the definition of propriety we refer to Section II and [10]. We briefly recapitulate these results on CWCU LMMSE estimation in this paper, and additionally derive the CWCU LMMSE estimator under the assumption of jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ (with no additional model assumptions). The main intent of this work, however, is the extension of the theoretical framework of CWCU linear estimation to CWCU widely linear estimators of the form

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{F}\mathbf{y}^* + \mathbf{b}, \tag{5}$$

with $\mathbf{E}$ and $\mathbf{F}$ as the estimator matrices. In general, when the measurement vector $\mathbf{y}$ turns improper [10], widely linear estimators are preferable over linear estimators [11]. For the LMMSE estimator and the widely linear MMSE (WLMMSE) estimator the particular form of the joint PDF $p(\mathbf{x}, \mathbf{y})$ does not play a role, the estimators are unambiguously defined by their first and second order statistics. However, the situation is different for CWCU estimators. The CWCU WLMMSE estimator always exists, and in the worst case it coincides with the best widely linear unbiased estimator (BWLUE). However, in a number of practically interesting situations, the CWCU WLMMSE estimator is able to outperform the BWLUE. In this paper we derive the CWCU WLMMSE estimator

(a) under the assumption of jointly generalized complex Gaussian $\mathbf{x}$ and $\mathbf{y}$,
(b) under the assumption of real $\mathbf{x}$, complex $\mathbf{y}$, and jointly Gaussian $\mathbf{x}$, $\mathrm{Re}\{\mathbf{y}\}$, and $\mathrm{Im}\{\mathbf{y}\}$.
(c) under the linear model assumption with generalized complex Gaussian $\mathbf{x}$ and zero mean noise with known second order statistics,
(d) under the linear model assumption with real Gaussian $\mathbf{x}$ and zero mean noise with known second order statistics,
(e) under the linear model assumption with mutually independent complex (and otherwise arbitrarily distributed) parameters and zero mean noise with known second order statistics, and
(f) under the linear model assumption with mutually independent real (and otherwise arbitrarily distributed) parameters and zero mean noise with known second order statistics.

We also address the analytical relationship between the CWCU WLMMSE and the WLMMSE estimator, which is not as straight forward as the relationship between the CWCU LMMSE and the LMMSE estimator regarded in [9].

The rest of the paper is organized as follows: In Section II we recapitulate the mathematical preliminaries required to derive the linear and particularly the widely linear estimators in this work. In Section III we extend linear CWCU estimation by a certain case not handled so far in our former papers. Then we turn to widely linear estimation. Section IV contains the prerequisites and derivations of the CWCU WLMMSE estimator under jointly Gaussian assumptions for $\mathbf{x}$ and $\mathbf{y}$. In Section V we assume an underlying linear model, which allows to relax some prerequisites from the previous section. Here we differ between correlated Gaussian parameter vectors and parameter vectors with mutually independent elements. Section VI contains an example, where the CWCU WLMMSE estimator is compared in performance to the well known estimators BLUE, BWLUE, LMMSE estimator, WLMMSE estimator and to the CWCU LMMSE estimator. Finally, Section VII compares all regarded estimators from an optimization point of view.

Notation:
Lower-case bold face variables ($\mathbf{a}$, $\mathbf{b}$,...) indicate vectors, and upper-case bold face variables ($\mathbf{A}$, $\mathbf{B}$,...) indicate matrices. We further use $\mathbb{R}$ and $\mathbb{C}$ to denote the set of real and complex numbers, respectively, $(\cdot)^T$ to denote transposition and $(\cdot)^H$ to denote conjugate transposition, $\mathbf{I}^{n\times n}$ to denote the identity matrix of size $n \times n$, and $\mathbf{0}^{m\times n}$ to denote the zero matrix of size $m \times n$. If the dimensions are clear from the context we simply write $\mathbf{I}$ and $\mathbf{0}$, respectively. $E[\cdot]$ denotes the expectation operator. In most of the cases we use an index to denote the averaging PDF, however, if the averaging PDF is clear from context, the index is sometimes omitted.

II. Preliminaries for Widely Linear Estimators

In this section we recapitulate the preliminaries required to derive the linear and particularly the widely linear estimators in this work. This section is more or less a shortened version of the corresponding parts in [10], where an excellent introduction to improper data and widely linear processing can be found.

A. Linear and Widely Linear Transformations

We start by constructing three closely related vectors from two real vectors $\mathbf{x}_r \in \mathbb{R}^n$ and $\mathbf{x}_i \in \mathbb{R}^n$. The first is the *real* *composite* $2n$-dimensional vector $\mathbf{x}_{\mathbb{R}} = \begin{bmatrix}\mathbf{x}_r^T & \mathbf{x}_i^T\end{bmatrix}^T$, obtained by stacking $\mathbf{x}_r$ on top of $\mathbf{x}_i$. The second is the *complex vector* $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i$, such that $\mathbf{x}_r = \mathrm{Re}\{\mathbf{x}\}$ and $\mathbf{x}_i = \mathrm{Im}\{\mathbf{x}\}$, and the third is the *complex augmented vector*

$$\underline{\mathbf{x}} = \begin{bmatrix}\mathbf{x}\\ \mathbf{x}^*\end{bmatrix}, \tag{6}$$

obtained by stacking $\mathbf{x}$ on top of its complex conjugate $\mathbf{x}^*$. The space of complex augmented vectors, whose bottom entries are the complex conjugates of the top entries, is denoted by $\mathbb{C}_*^{2n}$. Augmented vectors are always underlined. In much of our discussion, our focus will be on complex-valued quantities, where we will be using $\mathbf{x}$ and its augmentation $\underline{\mathbf{x}}$.

The complex augmented vector $\underline{\mathbf{x}} \in \mathbb{C}_*^{2n}$ is related to the real composite vector $\mathbf{x}_{\mathbb{R}} \in \mathbb{R}^{2n}$ as $\underline{\mathbf{x}} = \mathbf{T}_n\mathbf{x}_{\mathbb{R}}$ and $\mathbf{x}_{\mathbb{R}} = \frac{1}{2}\mathbf{T}_n^H\underline{\mathbf{x}}$, where the real-to-complex transformation matrix

$$\mathbf{T}_n = \begin{bmatrix}\mathbf{I} & j\mathbf{I}\\ \mathbf{I} & -j\mathbf{I}\end{bmatrix} \in \mathbb{C}^{2n\times 2n} \tag{7}$$

is unitary up to a factor of 2, i.e., $\mathbf{T}_n\mathbf{T}_n^H = \mathbf{T}_n^H\mathbf{T}_n = 2\mathbf{I}$. The complex augmented vector $\underline{\mathbf{x}}$ is obviously an equivalent redundant, but convenient, representation of $\mathbf{x}_{\mathbb{R}}$. When the size of $\mathbf{T}_n$ is clear, we may drop the subscript for economy.

In the following we consider widely linear transformations of the form

$$\mathbf{y} = \mathbf{H}_1\mathbf{x} + \mathbf{H}_2\mathbf{x}^*. \tag{8}$$

The augmented version of $\mathbf{y}$ can easily found to be

$$\underline{\mathbf{y}} = \begin{bmatrix}\mathbf{H}_1 & \mathbf{H}_2\\ \mathbf{H}_2^* & \mathbf{H}_1^*\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{x}^*\end{bmatrix} = \underline{\mathbf{H}}\underline{\mathbf{x}}. \tag{9}$$

The matrix $\underline{\mathbf{H}}$ is called an *augmented matrix*, it satisfies a particular block pattern, where the SE block is the conjugate of the NW block, and the SW block is the conjugate of the NE block. Obviously, the set of complex linear transformations $\mathbf{y} = \mathbf{H}_1\mathbf{x}$, with $\mathbf{H}_2 = \mathbf{0}$, or equivalently

$$\underline{\mathbf{y}} = \begin{bmatrix}\mathbf{H}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_1^*\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{x}^*\end{bmatrix} = \underline{\mathbf{H}}\underline{\mathbf{x}} \tag{10}$$

is a subset of the set of widely linear transformations.

B. Linear and Widely Linear Estimators

The estimators derived in this work will be compared to well known estimators like the BLUE, the BWLUE, the LMMSE and the WLMMSE estimator. Let $\mathbf{x} \in \mathbb{C}^n$ be the parameter vector to be estimated and $\mathbf{y} \in \mathbb{C}^m$ be the measurement vector, then a widely linear (or actually affine) estimator takes on the form

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{F}\mathbf{y}^* + \mathbf{b}. \tag{11}$$

In general widely linear estimators are superior to their linear counterparts as soon as the measurements $\mathbf{y}$ turn improper, see [12]- [20] for some possible applications of widely linear estimators. In the Sections on CWCU WLMMSE estimators we introduce

$$\mathbf{W} = \begin{bmatrix}\mathbf{E} & \mathbf{F}\end{bmatrix} \tag{12}$$

and write (11) usually in the form

$$\hat{\mathbf{x}} = \mathbf{W}\underline{\mathbf{y}} + \mathbf{b}. \tag{13}$$

Another way to express the estimator is its augmented version

$$\underline{\hat{\mathbf{x}}} = \begin{bmatrix}\mathbf{E} & \mathbf{F}\\ \mathbf{F}^* & \mathbf{E}^*\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \mathbf{y}^*\end{bmatrix} + \underline{\mathbf{b}} = \underline{\mathbf{E}}\underline{\mathbf{y}} + \underline{\mathbf{b}}. \tag{14}$$

For linear estimators we have $\mathbf{F} = \mathbf{0}$ such that $\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{b}$. The LMMSE estimator minimizing the BMSE cost function $E_{\mathbf{x},\mathbf{y}}[|\hat{x}_i - x_i|^2]$ for $i = 1, 2, \cdots, n$ and fulfilling the Bayesian unbiased constraint in (2) is given by

$$\hat{\mathbf{x}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{15}$$

It's widely linear counterpart, the WLMMSE estimator, is most compactly written in its augmented form [10], [11]

$$\underline{\hat{\mathbf{x}}} = E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]). \tag{16}$$

Many technical problems are described by the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{17}$$

where $\mathbf{H} \in \mathbb{C}^{m\times n}$ is a known observation matrix, $\mathbf{x}$ has mean $E_{\mathbf{x}}[\mathbf{x}]$ and covariance matrix $\mathbf{C}_{\mathbf{xx}}$, and $\mathbf{n} \in \mathbb{C}^m$ is a zero mean noise vector with covariance matrix $\mathbf{C}_{\mathbf{nn}}$ and independent of $\mathbf{x}$. The augmented version of (17) is

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}\underline{\mathbf{x}} + \underline{\mathbf{n}}, \tag{18}$$

where $\underline{\mathbf{H}}$ is defined as

$$\underline{\mathbf{H}} = \begin{bmatrix}\mathbf{H} & \mathbf{0}\\ \mathbf{0} & \mathbf{H}^*\end{bmatrix}. \tag{19}$$

If the parameter vector $\mathbf{x}$ and the measurement vector $\mathbf{y}$ are connected via the linear model, then the BLUE fulfilling the global unbiased constraint (1) is [21]

$$\hat{\mathbf{x}} = (\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}. \tag{20}$$

Its widely linear counterpart, the BWLUE, can be identified to be [10]

$$\underline{\hat{\mathbf{x}}} = (\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}})^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}}, \tag{21}$$

and it also fulfills (1). The BLUE and the BWLUE are usually treated in the classical framework, where $\mathbf{x}$ is assumed to be unknown but deterministic. The BWLUE is only able to outperform the BLUE if the noise $\mathbf{n}$ is improper (c.f. [10]).

C. Statistics of Complex-Valued Random Vectors

In order to characterize the second-order statistical properties of $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i$, we start by considering the real composite random vector $\mathbf{x}_{\mathbb{R}}$. Its covariance matrix is

$$\mathbf{C}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}} = E[(\mathbf{x}_{\mathbb{R}} - E[\mathbf{x}_{\mathbb{R}}])(\mathbf{x}_{\mathbb{R}} - E[\mathbf{x}_{\mathbb{R}}])^T] = \begin{bmatrix}\mathbf{C}_{\mathbf{x}_r\mathbf{x}_r} & \mathbf{C}_{\mathbf{x}_r\mathbf{x}_i}\\ \mathbf{C}_{\mathbf{x}_r\mathbf{x}_i}^T & \mathbf{C}_{\mathbf{x}_i\mathbf{x}_i}\end{bmatrix} \tag{22}$$

with $\mathbf{C}_{\mathbf{x}_r\mathbf{x}_r} = E[(\mathbf{x}_r - E[\mathbf{x}_r])(\mathbf{x}_r - E[\mathbf{x}_r])^T]$, $\mathbf{C}_{\mathbf{x}_r\mathbf{x}_i} = E[(\mathbf{x}_r - E[\mathbf{x}_r])(\mathbf{x}_i - E[\mathbf{x}_i])^T]$, and $\mathbf{C}_{\mathbf{x}_i\mathbf{x}_i} = E[(\mathbf{x}_i - E[\mathbf{x}_i])(\mathbf{x}_i - E[\mathbf{x}_i])^T]$. The augmented covariance matrix of $\mathbf{x}$ is

$$\underline{\mathbf{C}}_{\mathbf{xx}} = E[(\underline{\mathbf{x}} - E[\underline{\mathbf{x}}])(\underline{\mathbf{x}} - E[\underline{\mathbf{x}}])^H] \tag{23}$$

$$= \mathbf{T}\mathbf{C}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}}\mathbf{T}^H \tag{24}$$

$$= \begin{bmatrix}\mathbf{C}_{\mathbf{xx}} & \tilde{\mathbf{C}}_{\mathbf{xx}}\\ \tilde{\mathbf{C}}_{\mathbf{xx}}^* & \mathbf{C}_{\mathbf{xx}}^*\end{bmatrix} = \underline{\mathbf{C}}_{\mathbf{xx}}^H \in \mathbb{C}^{2n\times 2n}, \tag{25}$$

with $\mathbf{C}_{\mathbf{xx}} = E_{\mathbf{x}}[(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^H]$ as the (Hermitian and positive semi-definite) covariance matrix and $\tilde{\mathbf{C}}_{\mathbf{xx}} = E_{\mathbf{x}}[(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^T]$ as the complementary covariance matrix. For $\mathbf{C}_{\mathbf{xx}}$ and $\tilde{\mathbf{C}}_{\mathbf{xx}}$ we have

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}_r\mathbf{x}_r} + \mathbf{C}_{\mathbf{x}_i\mathbf{x}_i} + j(\mathbf{C}^T_{\mathbf{x}_r\mathbf{x}_i} - \mathbf{C}_{\mathbf{x}_r\mathbf{x}_i}) = \mathbf{C}^H_{\mathbf{xx}}, \quad (26)$$

and

$$\tilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}_r\mathbf{x}_r} - \mathbf{C}_{\mathbf{x}_i\mathbf{x}_i} + j(\mathbf{C}^T_{\mathbf{x}_r\mathbf{x}_i} + \mathbf{C}_{\mathbf{x}_r\mathbf{x}_i}) = \tilde{\mathbf{C}}^T_{\mathbf{xx}}, \quad (27)$$

respectively. $\tilde{\mathbf{C}}_{\mathbf{xx}}$ is sometimes also referred to as pseudo-covariance matrix or conjugate covariance matrix. If $\tilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{0}$, then the vector $\mathbf{x}$ is called *proper*, otherwise *improper* [22]-[27]. The conditions for propriety on the covariance and cross-covariance of real and imaginary parts $\mathbf{x}_r$ and $\mathbf{x}_i$ are $\mathbf{C}_{\mathbf{x}_r\mathbf{x}_r} = \mathbf{C}_{\mathbf{x}_i\mathbf{x}_i}$ and $\mathbf{C}_{\mathbf{x}_r\mathbf{x}_i} = -\mathbf{C}^T_{\mathbf{x}_r\mathbf{x}_i}$. When $x = x_r + jx_i$ is scalar, then $C_{x_r x_i} = 0$ is necessary for propriety. If $\mathbf{x}$ is proper, its Hermitian covariance matrix is

$$\mathbf{C}_{\mathbf{xx}} = 2\mathbf{C}_{\mathbf{x}_r\mathbf{x}_r} - 2j\mathbf{C}_{\mathbf{x}_r\mathbf{x}_i} = 2\mathbf{C}_{\mathbf{x}_i\mathbf{x}_i} + 2j\mathbf{C}^T_{\mathbf{x}_r\mathbf{x}_i}, \quad (28)$$

and its augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{xx}}$ is block-diagonal. If complex x is proper and scalar, then $\mathbf{C}_{xx} = 2\mathbf{C}_{x_r x_r} = 2\mathbf{C}_{x_i x_i}$. It is easy to see that propriety is preserved by strictly linear transformations, which are represented by block-diagonal augmented matrices.

D. Gaussian Random Vectors

To simplify notation we regard zero mean vectors in the following. Clearly the Gaussian PDF of the real composite $2n$-dimensional vector $\mathbf{x}_{\mathbb{R}} = \begin{bmatrix}\mathbf{x}_r^T & \mathbf{x}_i^T\end{bmatrix}^T$ is [10], [28]

$$p(\mathbf{x}_{\mathbb{R}}) = \frac{1}{(2\pi)^{\frac{2n}{2}}\sqrt{\det \mathbf{C}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}}}} \exp\left\{-\frac{1}{2}\mathbf{x}^T_{\mathbb{R}}\mathbf{C}^{-1}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}}\mathbf{x}_{\mathbb{R}}\right\}. \quad (29)$$

Using $\mathbf{x}_{\mathbb{R}} = \frac{1}{2}\mathbf{T}^H\underline{\mathbf{x}}$, $\mathbf{C}^{-1}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}} = \mathbf{T}^H\underline{\mathbf{C}}^{-1}_{\mathbf{xx}}\mathbf{T}$, and $\det \mathbf{C}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}} = 2^{-2n}\det \underline{\mathbf{C}}_{\mathbf{xx}}$, we obtain the PDF of complex $\mathbf{x}$ [29], [30]

$$p(\mathbf{x}) = \frac{1}{\pi^n\sqrt{\det \underline{\mathbf{C}}_{\mathbf{xx}}}} \exp\left\{-\frac{1}{2}\underline{\mathbf{x}}^H\underline{\mathbf{C}}^{-1}_{\mathbf{xx}}\underline{\mathbf{x}}\right\}. \quad (30)$$

This PDF depends algebraically on $\underline{\mathbf{x}}$, i.e., $\mathbf{x}$ and $\mathbf{x}^*$, but is interpreted as the joint PDF of $\mathbf{x}_r$ and $\mathbf{x}_i$, and can be used for proper or improper $\mathbf{x}$. In this work we call a complex vector $\mathbf{x}$ following this distribution *generalized complex Gaussian*. The simplification that occurs when $\tilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{0}$ is obvious and leads to the PDF of a complex proper Gaussian random vector $\mathbf{x}$:

$$p(\mathbf{x}) = \frac{1}{\pi^n \det \mathbf{C}_{\mathbf{xx}}} \exp\left\{-\mathbf{x}^H\mathbf{C}^{-1}_{\mathbf{xx}}\mathbf{x}\right\}. \quad (31)$$

III. CWCU LMMSE Estimation

We assume a vector parameter $\mathbf{x} \in \mathbb{C}^n$ is to be estimated based on a measurement vector $\mathbf{y} \in \mathbb{C}^m$. In the following we first derive the CWCU LMMSE estimator for jointly complex proper Gaussian $\mathbf{x}$ and $\mathbf{y}$, while no further assumptions on the measurement model are made. Subsequently, we briefly recapitulate the results from [9], where the CWCU LMMSE estimator has been derived for different linear model assumptions.

For jointly complex proper Gaussian $\mathbf{x}$ and $\mathbf{y}$, the optimum MMSE estimator is linear (or actually affine). In light of this we also constrain the CWCU estimator to be affine, such that

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{b}, \qquad \mathbf{E} \in \mathbb{C}^{n\times m}, \mathbf{b} \in \mathbb{C}^n. \quad (32)$$

Note that in LMMSE estimation no assumptions on the specific form of the joint PDF $p(\mathbf{x}, \mathbf{y})$ have to be made. However, the situation is different in CWCU LMMSE estimation. Let us consider the i^{th} component of the estimator

$$\hat{x}_i = \mathbf{e}_i^H\mathbf{y} + b_i, \quad (33)$$

where $\mathbf{e}_i^H$ denotes the i^{th} row of the estimator matrix $\mathbf{E}$. The conditional mean of $\hat{x}_i$ can be written as

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] + b_i. \quad (34)$$

A closer inspection of (34) reveals that $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ can be fulfilled for all possible x_i if the conditional mean $E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]$ is a linear (or actually affine) function of x_i, which is e.g. the case for jointly complex proper Gaussian $\mathbf{x}$ and $\mathbf{y}$.

A. CWCU LMMSE Estimation under the Jointly Gaussian Assumption

For proper and jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ the conditional mean $E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]$ is given by

$$E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] = E_{\mathbf{y}}[\mathbf{y}] + (\sigma^2_{x_i})^{-1}\mathbf{C}_{\mathbf{y}x_i}(x_i - E_{x_i}[x_i]), \quad (35)$$

where $\mathbf{C}_{\mathbf{y}x_i} = E_{\mathbf{y},x_i}[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])(x_i - E_{x_i}[x_i])^H]$, and $\sigma^2_{x_i}$ is the variance of x_i. $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{e}_i^H\mathbf{C}_{\mathbf{y}x_i} = \sigma^2_{x_i} \quad (36)$$

$$E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}] = b_i. \quad (37)$$

Inserting (33), (36) and (37) in the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ immediately leads to the constrained optimization problem

$$\mathbf{e}_{\mathrm{CL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H\mathbf{C}_{\mathbf{yy}}\mathbf{e}_i - \sigma^2_{x_i}\right) \quad \text{s.t. } \mathbf{e}_i^H\mathbf{C}_{\mathbf{y}x_i} = \sigma^2_{x_i}, \quad (38)$$

where "CL" shall stand for CWCU LMMSE. The solution can be found with the Lagrange multiplier method and is given by

$$\mathbf{e}^H_{\mathrm{CL},i} = \frac{\sigma^2_{x_i}}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}^{-1}_{\mathbf{yy}}\mathbf{C}_{\mathbf{y}x_i}}\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}^{-1}_{\mathbf{yy}}. \quad (39)$$

Using $\mathbf{E}_{\mathrm{CL}} = [e_{\mathrm{CL},1}, e_{\mathrm{CL},2}, \ldots, e_{\mathrm{CL},n}]^H$ together with (37) and (39) immediately leads us to the first part of the

Result 1. *If $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$ are jointly complex proper Gaussian then the CWCU LMMSE estimator minimizing the BMSEs $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ under the constraints $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ for $i = 1, 2, \cdots, n$ is given by*

$$\hat{\mathbf{x}}_{\mathrm{CL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CL}}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]), \quad (40)$$

with

$$\mathbf{E}_{\mathrm{CL}} = \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}^{-1}_{\mathbf{yy}}, \quad (41)$$

where the elements of the real diagonal matrix $\mathbf{D}$ are

$$[\mathbf{D}]_{i,i} = \frac{\sigma^2_{x_i}}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}^{-1}_{\mathbf{yy}}\mathbf{C}_{\mathbf{y}x_i}}. \quad (42)$$

The mean of the error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}_{\mathrm{CL}}$ *(in the Bayesian sense) is zero, and the error covariance matrix* $\mathbf{C}_{\mathbf{ee},\mathrm{CL}}$ *which is also the minimum BMSE matrix* $\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CL}}}$ *is*

$$\mathbf{C}_{\mathbf{ee},\mathrm{CL}} = \mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CL}}} = \mathbf{C}_{\mathbf{xx}} - \mathbf{AD} - \mathbf{DA} + \mathbf{DAD}, \tag{43}$$

with $\mathbf{A} = \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}$. *The minimum BMSEs are* $\mathrm{BMSE}(\hat{x}_{\mathrm{CL},i}) = [\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CL}}}]_{i,i} = \mathrm{MSE}(\hat{x}_{\mathrm{CL},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i)$ *and are given by*

$$\begin{aligned}\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i) &= E_{\hat{x}_{\mathrm{CL},i}|x_i}\left[|\hat{x}_{\mathrm{CL},i} - E_{\hat{x}_{\mathrm{CL},i}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i]|^2|x_i\right] \\ &= \mathbf{e}_{\mathrm{CL},i}^H\mathbf{C}_{\mathbf{yy}|x_i}\mathbf{e}_{\mathrm{CL},i} \qquad (44) \\ &= \frac{\left(\sigma_{x_i}^2\right)^2}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}} - \sigma_{x_i}^2 \qquad (45)\end{aligned}$$

with $\mathbf{C}_{\mathbf{yy}|x_i} = E_{\mathbf{y}|x_i}[(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i])(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i])^H|x_i]$.

The part on the error performance can simply be proved by inserting in the definitions of $\mathbf{e}$, $\mathbf{C}_{\mathbf{ee}}$, and $\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i)$, respectively. The conditional variance and the conditional MSE correspond since the conditional bias is zero. Furthermore, the Bayesian MSE and the conditional MSE correspond since the conditional MSE is independent of the parameter value x_i. From (41) it can be seen that the CWCU LMMSE estimator matrix can be derived as the product of the diagonal matrix $\mathbf{D}$ with the LMMSE estimator matrix $\mathbf{E}_{\mathrm{L}} = \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}$. Furthermore, we have $E_{\hat{x}_{\mathrm{L},i}|x_i}[\hat{x}_{\mathrm{L},i}|x_i] = [\mathbf{D}]_{i,i}^{-1}x_i + (1 - [\mathbf{D}]_{i,i}^{-1})E_{x_i}[x_i]$ for the LMMSE estimator. $\mathbf{D}$ can also be written as

$$\mathbf{D} = \mathrm{diag}\{\mathbf{C}_{\mathbf{xx}}\}\left(\mathrm{diag}\{\mathbf{A}\}\right)^{-1}. \tag{46}$$

B. CWCU LMMSE Estimation under Linear Model Assumptions

For completeness, the findings in [9] for CWCU LMMSE estimation under linear model assumptions will be stated in the following. Let $\mathbf{x}$ and $\mathbf{y}$ be connected via the linear model (17). Furthermore, let $\mathbf{h}_i \in \mathbb{C}^m$ be the i^{th} column of $\mathbf{H}$, $\bar{\mathbf{H}}_i \in \mathbb{C}^{m\times(n-1)}$ the matrix resulting from $\mathbf{H}$ by deleting $\mathbf{h}_i$, x_i be the i^{th} element of $\mathbf{x}$, and $\bar{\mathbf{x}}_i \in \mathbb{C}^{(n-1)}$ the vector resulting from $\mathbf{x}$ after deleting x_i. Then we can write

$$\mathbf{y} = \mathbf{h}_i x_i + \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n}, \tag{47}$$

and (33) becomes

$$\hat{x}_i = \mathbf{e}_i^H(\mathbf{h}_i x_i + \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n}) + b_i. \tag{48}$$

The conditional mean of $\hat{x}_i$ therefore is

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H\mathbf{h}_i x_i + \mathbf{e}_i^H\bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i] + b_i. \tag{49}$$

From (49) we can derive conditions that guarantee that the CWCU constraints (4) are fulfilled. There are at least the following possibilities:

1) (4) can be fulfilled for all possible x_i if the conditional mean $E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]$ is a linear function of x_i. For complex proper Gaussian $\mathbf{x}$ this condition holds (for all $i = 1, 2, ..., n$).
2) (4) can be fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i] = E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]$ for all possible x_i (and all $i = 1, 2, ..., n$), which is true if the elements x_i of $\mathbf{x}$ are mutually independent.
3) (4) is fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $\mathbf{e}_i^H\mathbf{h}_i = 1$ and $\mathbf{e}_i^H\bar{\mathbf{H}}_i = \mathbf{0}^T$ for $i = 1, 2, \cdots, n$, and if we set $b_i = 0$. These constraints and settings correspond to the ones of the BLUE.

We start with the first case, and recapitulate from [9]:

Result 2. *If the observed data* $\mathbf{y}$ *follow the linear model in* (17)*, where* $\mathbf{y} \in \mathbb{C}^m$ *is the data vector,* $\mathbf{H} \in \mathbb{C}^{m\times n}$ *is a known observation matrix,* $\mathbf{x} \in \mathbb{C}^n$ *is a parameter vector with prior complex proper Gaussian PDF* $\mathcal{CN}(E_{\mathbf{x}}[\mathbf{x}], \mathbf{C}_{\mathbf{xx}})$*, and* $\mathbf{n} \in \mathbb{C}^m$ *is a zero mean noise vector with covariance matrix* $\mathbf{C}_{\mathbf{nn}}$ *and independent of* $\mathbf{x}$ *(the PDF of* $\mathbf{n}$ *is otherwise arbitrary), then the CWCU LMMSE estimator minimizing the BMSEs* $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ *under the constraints* $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ *for* $i = 1, 2, \cdots, n$ *is given by* (40) *with*

$$\mathbf{E}_{\mathrm{CL}} = \mathbf{D}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}})^{-1}, \tag{50}$$

where the elements of the real diagonal matrix $\mathbf{D}$ *are*

$$[\mathbf{D}]_{i,i} = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{x}}\mathbf{H}^H(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}})^{-1}\mathbf{H}\mathbf{C}_{\mathbf{x}x_i}}. \tag{51}$$

Note that in Result 2 the requirements on $\mathbf{x}$ and $\mathbf{y}$ are weaker than in Result 1, since $\mathbf{x}$ and $\mathbf{y}$ need not to be jointly Gaussian. The PDF of $\mathbf{n}$ can in fact be arbitrary, the noise vector only has to be independent from $\mathbf{x}$. For the second case from above we recapitulate from [9]:

Result 3. *If the observed data* $\mathbf{y}$ *follow the linear model in* (17)*, where* $\mathbf{y} \in \mathbb{C}^m$ *is the data vector,* $\mathbf{H} \in \mathbb{C}^{m\times n}$ *is a known observation matrix,* $\mathbf{x} \in \mathbb{C}^n$ *is a parameter vector with mean* $E_{\mathbf{x}}[\mathbf{x}]$*, mutually independent elements and covariance matrix* $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \cdots, \sigma_{x_n}^2\}$*,* $\mathbf{n} \in \mathbb{C}^m$ *is a zero mean noise vector with covariance matrix* $\mathbf{C}_{\mathbf{nn}}$ *and independent of* $\mathbf{x}$ *(the PDF of* $\mathbf{n}$ *is otherwise arbitrary), then the CWCU LMMSE estimator minimizing the Bayesian MSEs* $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ *under the constraints* $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ *for* $i = 1, 2, \cdots, n$ *is given by* (40) *and* (50)*, where the elements of the real diagonal matrix* $\mathbf{D}$ *are*

$$[\mathbf{D}]_{i,i} = \frac{1}{\sigma_{x_i}^2\mathbf{h}_i^H(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}})^{-1}\mathbf{h}_i}. \tag{52}$$

The CWCU LMMSE estimator will in general not commute over linear transformations, an exception is discussed in [8].

IV. CWCU WLMMSE Estimation under Jointly Gaussian Assumptions

In the following we will derive the best widely linear (or actually affine) estimator in a BMSE sense, which fulfills the CWCU constraints in (4). We will do so by minimizing the BMSE cost function under certain constraints. Again as in the previous section we derive the estimator under different model assumptions. Practical relevant model assumptions which allow to derive a CWCU WLMMSE estimator that in general outperforms the BWLUE are listed in Section I. In the following, for every list entry (a) to (f) the according CWCU

WLMMSE estimator will be derived, yielding **Result 4** to **Result 9**.

We begin with the jointly Gaussian assumptions for $\mathbf{y}$ and $\mathbf{x}$. We assume the widely linear estimator for x_i to be of the form

$$\hat{x}_i = \mathbf{e}_i^H\mathbf{y} + \mathbf{f}_i^H\mathbf{y}^* + b_i, \quad \text{for } i = 1,2,...,n. \tag{53}$$

Eq. (53) can also be written as

$$\hat{x}_i = \mathbf{w}_i^H\underline{\mathbf{y}} + b_i, \quad \text{for } i = 1,2,...,n, \tag{54}$$

where we used

$$\mathbf{w}_i^H = \begin{bmatrix}\mathbf{e}_i^H & \mathbf{f}_i^H\end{bmatrix}. \tag{55}$$

The conditional mean of the estimator in (54) follows to

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] + b_i. \tag{56}$$

We first assume that $\mathbf{y}$ and $\mathbf{x}$ are generalized jointly Gaussian. Consequently, $E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ is linear in $\underline{x}_i$, namely

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{x}_i - E_{x_i}[\underline{x}_i]). \tag{57}$$

This leads to

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H\left(E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{x}_i - E_{x_i}[\underline{x}_i])\right) + b_i. \tag{58}$$

By setting (58) equal to $x_i = \begin{bmatrix}1 & 0\end{bmatrix}\underline{x}_i$ we find that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1} = \begin{bmatrix}1 & 0\end{bmatrix} \tag{59}$$

$$E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] = b_i. \tag{60}$$

These are the two conditions the widely linear estimator in (54) has to fulfill in order to become a CWCU estimator. For the derivation of the CWCU WLMMSE estimator we consider the BMSE cost function which follows to

$$\begin{aligned} J =& E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2] \\ =& E_{\mathbf{y},\mathbf{x}}[|\mathbf{w}_i^H\underline{\mathbf{y}} + b_i - x_i|^2] \\ =& E_{\mathbf{y},\mathbf{x}}[|\mathbf{w}_i^H(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - (x_i - E_{x_i}[x_i])|^2] \\ =& E_{\mathbf{y},\mathbf{x}}[|\mathbf{w}_i^H(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - \begin{bmatrix}1 & 0\end{bmatrix}(\underline{x}_i - E_{x_i}[\underline{x}_i])|^2] \\ =& \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix} - \\ & \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\mathbf{w}_i + \underbrace{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}}_{\sigma_{x_i}^2}. \end{aligned} \tag{61}$$

This result can be simplified by using (59), leading to the final optimization problem

$$\mathbf{w}_{\mathrm{CWL},i} = \arg\min_{\mathbf{w}_i}\left(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2\right) \quad \text{s.t. } \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1} = \begin{bmatrix}1 & 0\end{bmatrix}, \tag{62}$$

where "CWL" shall stand for CWCU WLMMSE. The solution of this optimization problem is derived in appendix A where we used the Lagrange multiplier method. The results of appendix A are summarized in the first part of

Result 4. *If* $\mathbf{x} \in \mathbb{C}^n$ *is a* ***complex valued parameter vector*** *and* $\mathbf{x}$ *and* $\mathbf{y} \in \mathbb{C}^m$ *are* ***generalized jointly Gaussian*** *then the CWCU WLMMSE estimator minimizing the BMSEs* $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ *under the constraints* $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ *for* $i = 1,2,\cdots,n$ *is*

$$\hat{\mathbf{x}}_{\mathrm{CWL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{W}_{\mathrm{CWL}}(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]), \tag{63}$$

with

$$\mathbf{W}_{\mathrm{CWL}} = \begin{bmatrix}\mathbf{w}_{\mathrm{CWL},1} & \mathbf{w}_{\mathrm{CWL},2} & \cdots & \mathbf{w}_{\mathrm{CWL},n}\end{bmatrix}^H, \tag{64}$$

where the rows of $\mathbf{W}_{\mathrm{CWL}}$ *are given by*

$$\mathbf{w}_{\mathrm{CWL},i}^H = \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{65}$$

The mean of the error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}_{\mathrm{CWL}}$ *(in the Bayesian sense) is zero, and the error covariance matrix* $\mathbf{C}_{\mathbf{ee},\mathrm{CWL}}$, *which is also the minimum BMSE matrix* $\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CWL}}}$, *is*

$$\begin{aligned}\mathbf{C}_{\mathbf{ee},\mathrm{CWL}} =& \mathbf{C}_{\mathbf{xx}} - \mathbf{W}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yx}}\begin{bmatrix}\mathbf{I}^{n\times n}\\ \mathbf{0}^{n\times n}\end{bmatrix} - \\ & \begin{bmatrix}\mathbf{I}^{n\times n} & \mathbf{0}^{n\times n}\end{bmatrix}\underline{\mathbf{C}}_{\mathbf{xy}}\mathbf{W}_{\mathrm{CWL}}^H + \\ & \mathbf{W}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{W}_{\mathrm{CWL}}^H. \end{aligned} \tag{66}$$

The minimum BMSEs are $\mathrm{BMSE}(\hat{x}_{\mathrm{CWL},i}) = [\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CWL}}}]_{i,i} = \mathrm{MSE}(\hat{x}_{\mathrm{CWL},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i)$ *and are given by*

$$\begin{aligned}\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) &= E[|\hat{x}_{\mathrm{CWL},i} - E[\hat{x}_{\mathrm{CWL},i}|x_i]|^2|x_i] \\ &= \mathbf{w}_{\mathrm{CWL},i}^H\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\mathbf{w}_{\mathrm{CWL},i}. \end{aligned} \tag{67}$$

$$\begin{aligned} &= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} \\ &\quad - \sigma_{x_i}^2. \end{aligned} \tag{68}$$

The part on the error performance can simply be proved by inserting in the definition of $\mathbf{e}$ and $\mathbf{C}_{\mathbf{ee}}$, respectively. The derivation of the conditional variance can be found in appendix B.

The CWCU WLMMSE estimator matrix $\mathbf{W}_{\mathrm{CWL}}$ from Result 4 can be derived from the WLMMSE estimator matrix $\underline{\mathbf{E}}_{\mathrm{WL}} = \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}$ according to

$$\mathbf{W}_{\mathrm{CWL}} = \begin{bmatrix}\mathbf{D}_1 & \mathbf{D}_2\end{bmatrix}\underline{\mathbf{E}}_{\mathrm{WL}}, \tag{69}$$

where the elements of the two diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are given by

$$[\mathbf{D}_1]_{i,i} = \left[\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1}\right]_{1,1}, \tag{70}$$

$$[\mathbf{D}_2]_{i,i} = \left[\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1}\right]_{1,2}. \tag{71}$$

In the next step we assume $\mathbf{x}$ to be a real valued vector, while $\mathbf{y}$ shall still be complex valued. In that case $\mathbf{y}$ and $\mathbf{x}$ are no longer generalized jointly Gaussian since the joint augmented covariance matrix is no longer invertible. Also $\underline{\mathbf{C}}_{x_ix_i}$ is not invertible, which was required in the derivation of Result 4, since

$$\underline{\mathbf{C}}_{x_ix_i} = \begin{bmatrix}\sigma_{x_i}^2 & \sigma_{x_i}^2\\ \sigma_{x_i}^2 & \sigma_{x_i}^2\end{bmatrix}. \tag{72}$$

However, we now assume the real composite vector

$$\mathbf{y}_{\mathbb{R}} = \begin{bmatrix}\mathrm{Re}\{\mathbf{y}\}\\ \mathrm{Im}\{\mathbf{y}\}\end{bmatrix} \in \mathbb{R}^{2m}, \tag{73}$$

and the real vector $\mathbf{x}$ to be jointly Gaussian. Hence, the conditional mean vector $E_{\mathbf{y}_\mathbb{R}|x_i}[\mathbf{y}_\mathbb{R}|x_i]$ is given by

$$E_{\mathbf{y}_\mathbb{R}|x_i}[\mathbf{y}_\mathbb{R}|x_i] = E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] + \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \quad (74)$$

By multiplying (74) with $\mathbf{T}$ from the left we obtain an expression for $E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ according to

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \quad (75)$$

With (75) the conditional mean of the estimator in (54) becomes

$$\begin{aligned}E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] =& \mathbf{w}_i^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] + b_i\\ =& \mathbf{w}_i^H\left(E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i])\right)\\ &+ b_i. \end{aligned} \quad (76)$$

By setting (76) equal to x_i we learn that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2} = 1 \quad (77)$$

$$E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] = b_i. \quad (78)$$

Simplifying the BMSE cost function in (61) using the constraint in (77) leads to the optimization problem

$$\begin{aligned}\mathbf{w}_{\mathrm{CWL},i} = \arg\min_{\mathbf{w}_i} &\left(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2\right)\\ &\text{s.t. } \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2} = 1.\end{aligned} \quad (79)$$

The derivation of this estimator can be found in appendix C, the results are summarized in the first part of

Result 5. *If $\mathbf{x} \in \mathbb{R}^n$ is a **real valued parameter vector** and $\mathbf{x}$ and $\mathbf{y}_\mathbb{R} \in \mathbb{C}^{2m}$ are **jointly Gaussian** then the CWCU WLMMSE estimator minimizing the BMSEs $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ under the constraints $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ for $i = 1, 2, \cdots, n$ is given by* (63) *where the rows of $\mathbf{W}_{\mathrm{CWL}}$ are given by*

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \quad (80)$$

The minimum BMSEs are $\mathrm{BMSE}(\hat{x}_{\mathrm{CWL},i}) = [\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CWL}}}]_{i,i} = \mathrm{MSE}(\hat{x}_{\mathrm{CWL},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i)$ *and are given by*

$$\begin{aligned}\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) =& \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\mathbf{w}_i\\ =& \frac{\left(\sigma_{x_i}^2\right)^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}} - \sigma_{x_i}^2.\end{aligned} \quad (81)$$

The derivation of the conditional variances can be found in appendix D. An alternative representation of (80) can be obtained by utilizing $\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}} = \mathbf{C}_{x_i\underline{\mathbf{y}}}$, yielding

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{C}_{\underline{\mathbf{y}}x_i}}\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \quad (82)$$

The CWCU WLMMSE estimator matrix $\mathbf{W}_{\mathrm{CWL}}$ from Result 5 can be derived from the WLMMSE estimator matrix $\underline{\mathbf{E}}_{\mathrm{WL}} = \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}$ according to

$$\mathbf{E}_{\mathrm{CWL}} = \mathbf{D}\begin{bmatrix}\mathbf{I}^{n\times n} & \mathbf{0}^{n\times n}\end{bmatrix}\underline{\mathbf{E}}_{\mathrm{WL}}, \quad (83)$$

where the elements of the diagonal matrix $\mathbf{D}$ are given by

$$[\mathbf{D}]_{i,i} = \frac{\sigma_{x_i}^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}. \quad (84)$$

Note that this estimator always yields real values since $\mathbf{F}_{\mathrm{CWL}} = \mathbf{E}_{\mathrm{CWL}}^*$ or $\mathbf{W}_{\mathrm{CWL}} = \begin{bmatrix}\mathbf{E}_{\mathrm{CWL}} & \mathbf{F}_{\mathrm{CWL}}\end{bmatrix} = \begin{bmatrix}\mathbf{E}_{\mathrm{CWL}} & \mathbf{E}_{\mathrm{CWL}}^*\end{bmatrix}$.

V. CWCU WLMMSE Estimation under Linear Model Assumptions

In the following it will be seen that some of the prerequisites of Result 4 and 5 can be relaxed when incorporating details of the data model into the derivation of the estimator. From now on we limit our considerations to the linear model in (17). Statistical assumptions on $\mathbf{x}$ and $\mathbf{n}$ will vary in the following.

For the linear model the augmented covariance matrices required in (65), (66), (80) and (81) become

$$\underline{\mathbf{C}}_{x_i\mathbf{y}} = \underline{\mathbf{C}}_{x_i\mathbf{x}}\underline{\mathbf{H}}^H \quad (85)$$

$$\underline{\mathbf{C}}_{\mathbf{y}x_i} = \underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{x}x_i} \quad (86)$$

$$\underline{\mathbf{C}}_{\mathbf{yy}} = \underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H + \underline{\mathbf{C}}_{\mathbf{nn}} \quad (87)$$

$$\underline{\mathbf{C}}_{\mathbf{xy}} = \underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H. \quad (88)$$

If the assumptions made on the linear model above hold and if $\mathbf{x}$ and $\mathbf{n}$ are both generalized complex Gaussian, then they are generalized jointly Gaussian. Furthermore, since $[\underline{\mathbf{x}}^T, \underline{\mathbf{y}}^T]^T$ is a linear transformation of $[\underline{\mathbf{x}}^T, \underline{\mathbf{n}}^T]^T$, $\mathbf{x}$ and $\mathbf{y}$ are generalized jointly Gaussian, too. We could therefore simply insert (85)-(88) into the equations given in Result 4. However, the jointly Gaussian assumption for $\mathbf{x}$ and $\mathbf{n}$ can significantly be relaxed. This can be shown by incorporating the linear model assumption already earlier in the derivation of the estimator. With the notation

$$\underline{\mathbf{H}}_i = \begin{bmatrix}\mathbf{h}_i & \mathbf{0}\\ \mathbf{0} & \mathbf{h}_i^*\end{bmatrix} \in \mathbb{C}^{2m\times 2},\ \underline{\bar{\mathbf{H}}}_i = \begin{bmatrix}\bar{\mathbf{H}}_i & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{H}}_i^*\end{bmatrix} \in \mathbb{C}^{2m\times(2n-2)} \quad (89)$$

the augmented form of (47) follows to

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}_i\underline{x}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}}. \quad (90)$$

Incorporating (90) into the conditional mean of the estimator $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i]$ yields

$$\begin{aligned}E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] =& E_{\mathbf{y}|x_i}[\mathbf{w}_i^H\underline{\mathbf{y}} + b_i|x_i]\\ =& E_{\mathbf{n},\bar{\mathbf{x}}_i|x_i}[\mathbf{w}_i^H(\underline{\mathbf{H}}_i\underline{x}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}}) + b_i|x_i]\\ =& \mathbf{w}_i^H(\underline{\mathbf{H}}_i\underline{x}_i + \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]) + b_i.\end{aligned} \quad (91)$$

From (91) we can derive conditions that guarantee that the CWCU constraints (4) are fulfilled. There are at least the following possibilities:

1) (4) can be fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $\mathbf{x}$ is generalized complex Gaussian (the reasoning follows immediately).

2) (4) can be fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i] = E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i]$ for all possible x_i (and all $i = 1, 2, ..., n$), which is true if the elements x_i of $\mathbf{x}$ are mutually independent.
3) (4) is fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $\mathbf{w}_i^H \underline{\mathbf{H}}_i = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $\mathbf{w}_i^H \underline{\bar{\mathbf{H}}}_i = \mathbf{0}^T$ for $i = 1, 2, \cdots, n$, and if we set $b_i = 0$. These constraints and settings correspond to the ones of the BWLUE [10].

A. Correlated Gaussian Parameters

We start with the first case from above, assume a generalized complex Gaussian parameter vector $\mathbf{x}$, and begin the derivation of the i^{th} component $\hat{x}_i$ of the estimator. Because of the Gaussian assumption we have

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H \Big(\underline{\mathbf{H}}_i \underline{x}_i + \underline{\bar{\mathbf{H}}}_i \big(E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] + \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} (\underline{x}_i - E_{x_i}[\underline{x}_i]) \big) \Big) + b_i. \quad (92)$$

By setting (92) equal to $x_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{x}_i$ one can see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H \underline{\mathbf{H}}_i + \mathbf{w}_i^H \underline{\bar{\mathbf{H}}}_i \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad (93)$$

$$b_i = -\mathbf{w}_i^H \underline{\bar{\mathbf{H}}}_i \left(E_{\bar{\mathbf{x}}}[\underline{\bar{\mathbf{x}}}_i] - \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} E_{x_i}[\underline{x}_i] \right). \quad (94)$$

After some algebraic manipulations (93) and (94) can compactly be written as

$$\mathbf{w}_i^H \underline{\mathbf{H}} \underline{\mathbf{C}}_{\mathbf{x} x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad (95)$$

$$b_i = E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. \quad (96)$$

Eq. (95) could also have been derived from (59) by assuming an underlying linear model. However, the approach in this section shows that the noise need not to be Gaussian. Inserting into the BMSE cost function leads to the optimization problem

$$\mathbf{w}_{\mathrm{CWL},i} = \arg\min_{\mathbf{w}_i} \left(\mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{w}_i - \sigma_{x_i}^2 \right) \quad \text{s.t. } \mathbf{w}_i^H \underline{\mathbf{H}} \underline{\mathbf{C}}_{\mathbf{x} x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix}. \quad (97)$$

The solution to this constrained optimization problem can be found using the Lagrange multiplier method and is given in

Result 6. *If the observed data $\mathbf{y}$ follow the linear model in* (17)*, where $\mathbf{y} \in \mathbb{C}^m$ is the data vector, $\mathbf{H} \in \mathbb{C}^{m \times n}$ is a known observation matrix, $\mathbf{x} \in \mathbb{C}^n$ is a **generalized complex Gaussian parameter vector** with mean vector $E_{\mathbf{x}}[\mathbf{x}]$ and augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{xx}}$, and $\mathbf{n} \in \mathbb{C}^m$ is a zero mean noise vector with augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$ and independent of $\mathbf{x}$ (the PDF of $\mathbf{n}$ is otherwise arbitrary), then the CWCU WLMMSE estimator minimizing the BMSEs $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ under the constraints $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ for $i = 1, 2, \cdots, n$ is given by* (63)-(65) *with* (85)-(88) *inserted for the augmented covariance matrices.*

Equivalent to Result 4, Result 6 cannot be applied if $\mathbf{x}$ is real valued since then $\underline{\mathbf{C}}_{x_i x_i}$ is not invertible. By following similar steps as in Section IV one can show that for real valued Gaussian parameter vectors the following optimization problem occurs:

$$\mathbf{w}_{\mathrm{CWL},i} = \arg\min_{\mathbf{w}_i} \left(\mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{w}_i - \sigma_{x_i}^2 \right) \quad \text{s.t. } \mathbf{w}_i^H \underline{\mathbf{H}} \mathbf{C}_{\mathbf{x} x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\sigma_{x_i}^2} = 1. \quad (98)$$

The solution is given in

Result 7. *If the observed data $\mathbf{y}$ follow the linear model in* (17)*, where $\mathbf{y} \in \mathbb{C}^m$ is the data vector, $\mathbf{H} \in \mathbb{C}^{m \times n}$ is a known observation matrix, $\mathbf{x} \in \mathbb{R}^n$ is a **real valued Gaussian parameter vector** with prior PDF $\mathcal{N}(E_{\mathbf{x}}[\mathbf{x}], \mathbf{C}_{\mathbf{xx}})$, and $\mathbf{n} \in \mathbb{C}^m$ is a zero mean noise vector with augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$ and independent of $\mathbf{x}$ (the PDF of $\mathbf{n}$ is otherwise arbitrary), then the CWCU WLMMSE estimator minimizing the BMSEs $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ under the constraints $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ for $i = 1, 2, \cdots, n$ is given by* (63)*,* (64) *and* (80) *with* (85)-(88) *inserted for the augmented covariance matrices.*

B. Mutually Independent Parameters

For mutually independent parameters it is possible to further relax the prerequisites on $\mathbf{x}$. In this case (91) becomes

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H \underline{\mathbf{H}}_i \underline{x}_i + \mathbf{w}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] + b_i, \quad (99)$$

since $E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]$ is no longer dependent on x_i. Let $\mathbf{x}$ be complex. By setting (99) equal to $x_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{x}_i$ we see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H \underline{\mathbf{H}}_i = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad (100)$$

$$b_i = -\mathbf{w}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] = E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. \quad (101)$$

Eq. (100) could also have been derived from (93) by assuming the elements of $\mathbf{x}$ to be mutually independent. However, the approach in this section shows that no further assumptions (like the Gaussian assumption) on the PDF of $\mathbf{x}$ have to be made in the case of mutually independent parameters. Inserting in the BMSE cost function and simplifying leads to the optimization problem

$$\mathbf{w}_{\mathrm{CWL},i} = \arg\min_{\mathbf{w}_i} \left(\mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{w}_i - \sigma_{x_i}^2 \right) \quad \text{s.t. } \mathbf{w}_i^H \underline{\mathbf{H}}_i = \begin{bmatrix} 1 & 0 \end{bmatrix}. \quad (102)$$

Solving this constrained optimization problem leads to

Result 8. *If the observed data $\mathbf{y}$ follow the linear model in* (17)*, where $\mathbf{y} \in \mathbb{C}^m$ is the data vector, $\mathbf{H} \in \mathbb{C}^{m \times n}$ is a known observation matrix, $\mathbf{x} \in \mathbb{C}^n$ is a complex parameter vector with mean $E_{\mathbf{x}}[\mathbf{x}]$ and **mutually independent elements** such that $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \cdots, \sigma_{x_n}^2\}$ and $\tilde{\mathbf{C}}_{\mathbf{xx}} = \mathrm{diag}\{\tilde{\sigma}_{x_1}^2, \tilde{\sigma}_{x_2}^2, \cdots, \tilde{\sigma}_{x_n}^2\}$, $\mathbf{n} \in \mathbb{C}^m$ is a zero mean noise vector with augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$ and independent of $\mathbf{x}$ (the PDF of $\mathbf{x}$ and $\mathbf{n}$ are otherwise arbitrary). Then the CWCU WLMMSE estimator minimizing the BMSEs $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ under the constraints $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ for $i = 1, 2, \cdots, n$ is given by* (63) *and* (64)*, where the rows of $\mathbf{W}_{\mathrm{CWL}}$ are given by*

$$\mathbf{w}_{\mathrm{CWL},i}^H = \begin{bmatrix} 1 & 0 \end{bmatrix} (\underline{\mathbf{H}}_i^H \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{H}}_i)^{-1} \underline{\mathbf{H}}_i^H \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}, \quad (103)$$

and $\underline{\mathbf{C}}_{\mathbf{yy}}$ *is defined in* (87).

By applying Result 6, and using $\underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{x}x_i} = \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_ix_i}$, which holds for mutually independent parameters, we also arrive at the same formulas as in Result 8, however, the prerequisites in Result 8 are much more relaxed. We note that the constraint $\mathbf{w}_i^H\underline{\mathbf{H}}_i = \begin{bmatrix}1 & 0\end{bmatrix}$ is equivalent to

$$\mathbf{f}_i^H\mathbf{h}_i = 1, \qquad \mathbf{g}_i^H\mathbf{h}_i^* = 0. \tag{104}$$

We now again turn to the case that $\mathbf{x}$ is a real valued parameter vector. In that case we can further relax the constraints: By using $\underline{x}_i = \begin{bmatrix}x_i & x_i\end{bmatrix}^T$ (which is true for real x_i) in (99) we immediately see that for the real case the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{f}_i^H\mathbf{h}_i + \mathbf{g}_i^H\mathbf{h}_i^* = \mathbf{w}_i^H\underline{\mathbf{H}}_i\begin{bmatrix}1\\1\end{bmatrix} = 1, \tag{105}$$

which is a softer requirement than $\mathbf{w}_i^H\underline{\mathbf{H}}_i = \begin{bmatrix}1 & 0\end{bmatrix}$ (compare (105) with (104)). Inserting in the BMSE cost function and simplifying leads to the optimization problem

$$\begin{aligned}\mathbf{w}_{\mathrm{CWL},i} = \arg\min_{\mathbf{w}_i} \ &\left(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2\right)\\ \text{s.t. } &\mathbf{w}_i^H\underline{\mathbf{H}}_i\begin{bmatrix}1\\1\end{bmatrix} = 1.\end{aligned} \tag{106}$$

The CWCU WLMMSE estimator for the real parameter case is the solution to this optimization problem and is given in

Result 9. *If the observed data* $\mathbf{y}$ *follow the linear model in* (17)*, where* $\mathbf{y} \in \mathbb{C}^m$ *is the data vector,* $\mathbf{H} \in \mathbb{C}^{m\times n}$ *is a known observation matrix,* $\mathbf{x} \in \mathbb{R}^n$ *is a* ***real valued parameter vector*** *with mean* $E_{\mathbf{x}}[\mathbf{x}]$*,* ***mutually independent elements*** *and covariance matrix* $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \cdots, \sigma_{x_n}^2\}$*,* $\mathbf{n} \in \mathbb{C}^m$ *is a zero mean noise vector with augmented covariance matrix* $\underline{\mathbf{C}}_{\mathbf{nn}}$ *and independent of* $\mathbf{x}$ *(the joint PDF of* $\mathbf{x}$ *and* $\mathbf{n}$ *is otherwise arbitrary). Then the CWCU WLMMSE estimator minimizing the BMSEs* $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ *under the constraints* $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ *for* $i = 1, 2, \cdots, n$ *is given by* (63) *and* (64)*, where the rows of* $\mathbf{W}_{\mathrm{CWL}}$ *are given by*

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{1}{\begin{bmatrix}1 & 1\end{bmatrix}\underline{\mathbf{H}}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{H}}_i\begin{bmatrix}1\\1\end{bmatrix}}\begin{bmatrix}1 & 1\end{bmatrix}\underline{\mathbf{H}}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{107}$$

By using the augmented vector $\underline{\mathbf{h}}_i = \begin{bmatrix}\mathbf{h}_i^T & \mathbf{h}_i^H\end{bmatrix}^T$, (107) simplifies to

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{1}{\underline{\mathbf{h}}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{h}}_i}\underline{\mathbf{h}}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{108}$$

Let's recall that the CWCU WLMMSE estimator from Result 4 applied on the linear model leads to the same formulas as in Result 6. However, the prerequisites in Result 4 and 6 differ. The same statement holds for Result 5 and 7, Result 6 and 8, and Result 7 and 9, respectively.

C. Separated Real and Imaginary Parts

Another way to estimate a complex parameter vector is to rewrite the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ according to

$$\mathbf{y} = \underbrace{\begin{bmatrix}\mathbf{H} & i\mathbf{H}\end{bmatrix}}_{\mathbf{H}' \in \mathbb{C}^{m\times 2n}}\underbrace{\begin{bmatrix}\mathrm{Re}\{\mathbf{x}\}\\ \mathrm{Im}\{\mathbf{x}\}\end{bmatrix}}_{\mathbf{x}_{\mathbb{R}} \in \mathbb{C}^{2n}} + \mathbf{n}, \tag{109}$$

and estimate the real and imaginary parts of the parameter vector separately. With (109), the parameter vector is real valued which enables us to use the CWCU WLMMSE estimator for real valued parameter vectors. The estimated real and imaginary parts can then be combined to a complex estimator for the parameter vector $\mathbf{x}$. It is to note that this estimator is in general not a CWCU estimator for the complex parameters x_i, but it is a CWCU estimator for $\mathrm{Re}\{x_i\}$ and $\mathrm{Im}\{x_i\}$, since we forced $E[\widehat{\mathrm{Re}\{x_i\}}|\mathrm{Re}\{x_i\}] = \mathrm{Re}\{x_i\}$ and $E[\widehat{\mathrm{Im}\{x_i\}}|\mathrm{Im}\{x_i\}] = \mathrm{Im}\{x_i\}$ for $i = 1, 2, \cdots, n$. This is why this estimator will be denoted as part-wise conditionally unbiased WLMMSE (PWCU WLMMSE) estimator. In general this estimator features a lower BMSE compared to its CWCU counterpart, since conditioning separately on the real and on the imaginary parts are in general weaker constraints than conditioning on the complex parameters. However, there exist cases where CWCU and PWCU estimators feature the same BMSE performance.

VI. Example: Widely Linear Estimation of Improper Data

For examples on the application of CWCU LMMSE estimators for proper parameter vectors we refer the reader to [8] and [9], where we investigated channel and data estimation applications. Here we test the CWCU WLMMSE estimator, and compare it with the BLUE, the LMMSE estimator, the CWCU LMMSE estimator, the BWLUE, and the WLMMSE estimator. We do this by estimating a complex constant and the complex amplitude of a complex exponential in the presence of noise. The signal model is $y[k] = x_1 + 1.5x_2e^{j6k} + n[k]$ for $k = 0, 1, \cdots, 5$. Which can easily be brought to the form of a linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$. We assume the noise vector $\mathbf{n}$ to be complex proper Gaussian $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{nn}})$ with

$$\mathbf{C}_{\mathbf{nn}} = \mathrm{diag}\{0.1,\ 0.06,\ 0.3,\ 0.2,\ 0.15,\ 0.1\}. \tag{110}$$

Furthermore, in our experiment we let the covariance matrices of the real and imaginary parts of $\mathbf{x}$ and the cross-covariance matrix be

$$\mathbf{C}_{\mathrm{Re}\{\mathbf{x}\}\mathrm{Re}\{\mathbf{x}\}} = \mathrm{diag}\{1,\ 0.6\} \tag{111}$$

$$\mathbf{C}_{\mathrm{Im}\{\mathbf{x}\}\mathrm{Im}\{\mathbf{x}\}} = k\,\mathrm{diag}\{1,\ 0.6\} \tag{112}$$

$$\mathbf{C}_{\mathrm{Re}\{\mathbf{x}\}\mathrm{Im}\{\mathbf{x}\}} = \mathbf{0}^{2\times 2}, \tag{113}$$

where the scalar k in $\mathbf{C}_{\mathrm{Im}\{\mathbf{x}\}\mathrm{Im}\{\mathbf{x}\}}$ can vary between 10^{-4} and 10^2. According to this setup the parameter vector $\mathbf{x}$ is improper for $k \neq 1$ and proper for $k = 1$. We start with $k = 10^{-4}$, such that the parameter vector is close to real, and test all the estimators listed in Table I. Then we increase k stepwise, such that the imaginary part of $\mathbf{x}$ becomes more and more significant, and repeat the estimation procedures accordingly.

TABLE I
ESTIMATORS USED FOR THE PROBLEM DESCRIBED IN SECTION VI

Estimator	***Section***	***Equation***
BLUE	II-B	(20)
LMMSE	II-B	(15)
CWCU LMMSE	III	Result 3
BWLUE	II-B	(21)
WLMMSE	II-B	(16)
CWCU WLMMSE	V-B	Result 8
CWCU WLMMSE for real parameter vectors	V-B	Result 9
PWCU WLMMSE	V-C	Result 9

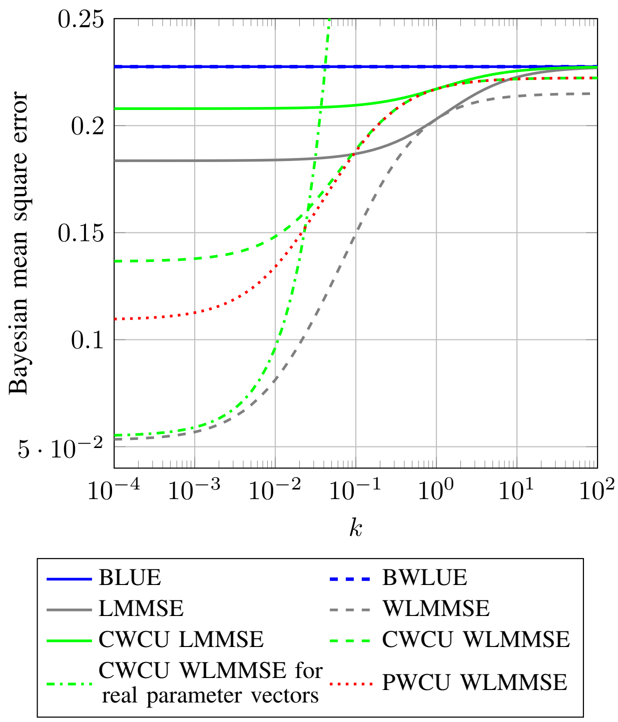

Fig. 2. BMSE values plotted over the scaling factor k which defines the variances of the imaginary parts. The variances of the real parts have been kept constant.

The result is a BMSE curve for each estimator in dependence of k. With this setup we can observe how the estimators perform for highly improper and also proper data within the scope of this example. Note that we also test the CWCU WLMMSE estimator for real parameter vectors. Clearly this estimator only perfectly fulfills the CWCU constraints once the parameter vector is in fact real. However, for $k = 10^{-4}$ it makes sense to apply this estimator since in that case the imaginary parts of the parameters are negligible compared to the real parts. Of course for increasing k the application of this estimator does not make sense.

Fig. 2 shows the resulting BMSE curves plotted over the scaling factor k. Clearly, the WLMMSE estimator features the best BMSE performance for all k since this estimator minimizes the BMSE cost function without any constraints. The BLUE and the BWLUE show the worst performance. They perform equal, which is clear since the BWLUE is only able to outperform the BLUE in case of improper noise. Both estimators show the same performance for all k, because they do not incorporate statistical knowledge on the parameters.

Especially for small k, which corresponds to highly improper data, the LMMSE estimator's performance is far below the one of the WLMMSE estimator, while for $k = 1$ (the proper case) they clearly perform equal. This impressively shows that the LMMSE estimator is not able to exploit information about the improperness of $\mathbf{x}$. The CWCU WLMMSE estimator derived in this work also significantly outperforms the LMMSE estimator for small values of k, and it is also in front for large $k > 10$. For $k = 10^{-4}$, where we approximately have a real valued parameter vector, the CWCU WLMMSE estimator for real parameter vectors comes quite close to the WLMMSE estimator. However, it is interesting to note that the CWCU WLMMSE estimator for complex parameter vectors does not converge to the CWCU WLMMSE estimator for real parameter vectors for $k \to -\infty$. Consequently, once we know from the application that the parameter vector is real we shall definitely apply the CWCU WLMMSE estimator for real parameter vectors. In this example it can also be seen that the PWCU WLMMSE estimator particularly outperforms the CWCU WLMMSE estimator for small k.

We already noted that for $k = 1$ (the proper case), the LMMSE estimator and the WLMMSE estimator perform equal, the same is true for the CWCU LMMSE and the CWCU WLMMSE estimator.

For $k \gg 1$, the variances of the imaginary parts of the parameters are way bigger than the noise variances. Hence, the prior knowledge about $\mathbf{C}_{\text{Im}\{\mathbf{x}\}\text{Im}\{\mathbf{x}\}}$ become less important. What's left is the prior knowledge about $\mathbf{C}_{\text{Re}\{\mathbf{x}\}\text{Re}\{\mathbf{x}\}}$. Linear estimators are not able to incorporate this particular knowledge, and they all converge towards the BLUE's performance for large k. The WLMMSE estimator and the CWCU WLMMSE estimator still keep a little performance gain compared to the linear estimators due to the incorporation of the prior knowledge about the improperness of $\mathbf{x}$.

To conclude this example we can state that the CWCU WLMMSE estimator significantly outperforms its globally unbiased counterparts BLUE and BWLUE, and compared to the WLMMSE estimator the CWCU WLMMSE estimator features the favorable property of component-wise conditionally unbiasedness.

VII. ESTIMATOR COMPARISON

In standard literature [1] the BLUE is treated as a classical linear estimator $\hat{\mathbf{x}} = \mathbf{E}\mathbf{y}$, which is derived by minimizing the estimators variance subject to the unbiased constraint $\mathbf{E}\mathbf{H} = \mathbf{I}$:

$$\min \text{var}(\hat{\mathbf{x}}) \quad \text{s.t. } \mathbf{E}\mathbf{H} = \mathbf{I}. \tag{114}$$

It can be shown that this estimator can also be derived in the Bayesian framework by minimizing the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ subject to the (global) unbiased constraint $\mathbf{E}\mathbf{H} = \mathbf{I}$, such that the BLUE can also be interpreted as a Bayesian estimator. Similar arguments also hold for the

TABLE II
LINEAR AND WIDELY LINEAR ESTIMATORS AND THEIR CONSTRAINTS

Estimator	***Constraints***
BLUE	$\mathbf{EH} = \mathbf{I}$, $\mathbf{F} = \mathbf{0}$
LMMSE	$\mathbf{F} = \mathbf{0}$
CWCU LMMSE	$\text{diag}\{\mathbf{EH}\} = \mathbf{1}$, $\mathbf{F} = \mathbf{0}$
BWLUE	$\mathbf{EH} = \mathbf{I}$, $\mathbf{FH}^* = \mathbf{0}$
WLMMSE	-
CWCU WLMMSE	$\text{diag}\{\mathbf{EH}\} = \mathbf{1}$, $\text{diag}\{\mathbf{FH}^*\} = \mathbf{0}$
CWCU WLMMSE for real parameter vectors	$\text{diag}\{\mathbf{EH}\} + \text{diag}\{\mathbf{FH}^*\} = \mathbf{1}$

BWLUE. Hence, every estimator regarded in this work can be derived by minimizing the BMSE cost function subject to particular constraints (except the WLMMSE estimator which minimizes the BMSE cost function without any constraint but the widely linear restriction). In the following we concentrate on the linear model case with a parameter vector having mutually independent parameters, furthermore we assume the parameter vector and the measurement vector to have zero mean. These assumptions are made since then also the constraints for the CWCU estimators take on quite simple forms (while the constraints on BLUE, BWLUE, LMMSE and WLMMSE estimator do not change by making particular assumptions on the PDF of $\mathbf{x}$). Let the general widely linear estimator for this setup be of the form

$$\hat{\mathbf{x}} = \mathbf{W}\underline{\mathbf{y}} = \begin{bmatrix}\mathbf{E} & \mathbf{F}\end{bmatrix} \underline{\mathbf{y}} = \mathbf{E}\mathbf{y} + \mathbf{F}\mathbf{y}^*. \tag{115}$$

Table II lists all the estimators regarded in this work together with the constraints that have to be fulfilled for this particular setup when minimizing the BMSE cost function. The estimator with the most stringent constraint, which is the BLUE, will in general perform worst in a BMSE sense. On the other hand, the BLUE produces unbiased estimates in the classical sense. The LMMSE estimator and the WLMMSE estimator, while performing better in a BMSE sense than the BLUE and the BWLUE, respectively, are conditionally biased, leading to effects demonstrated in Fig. 1. The CWCU estimators derived in this paper prevent this property, and in contrast to the BLUE and the BWLUE they are in general able to incorporate prior knowledge about the statistics of the parameter vector, which can lead to a significant performance gain over these classical estimators (c.f. Section VI).

VIII. Conclusion

In this paper we completed previous findings on CWCU LMMSE estimation and derived an analytical solution in dependence on the first and second order statistics for the case, that the parameters and measurements are jointly Gaussian. The main intent of the work, however, was the extension of component-wise conditionally unbiased estimation to widely linear estimators. We derived CWCU WLMMSE estimators for a number of different preconditions, and started with jointly Gaussian parameters and measurements. In the next step we made linear model assumptions, investigated the cases of jointly Gaussian and mutually independent parameters, and showed that the jointly Gaussian assumption of the parameter and measurement vectors can significantly be relaxed. In particular, the PDF of the noise can be arbitrary, and in case of mutually independent parameters their joint PDF can also be of any form. Furthermore, for every regarded set of preconditions we distinguished between improper complex and real parameters, which lead to different analytical expressions in each case.

Appendix A
Derivation of the CWCU WLMMSE estimator under the generalized jointly Gaussian assumption of $\mathbf{x}$ and $\mathbf{y}$

In appendix A we solve the optimization problem given in (62) using the Lagrange multiplier method. We start with the Lagrangian cost function which is

$$J' = \mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{w}_i - \sigma_{x_i}^2 + \lambda^H \left(\underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}} \mathbf{w}_i - \begin{bmatrix}1\\0\end{bmatrix} \right). \tag{116}$$

Using Wirtinger's calculus [31] for complex derivatives, we obtain

$$\frac{\partial J'}{\partial \mathbf{w}_i} = \underline{\mathbf{C}}_{\mathbf{yy}}^T \mathbf{w}_i^* + (\lambda^H \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}})^T. \tag{117}$$

By setting (117) equal to zero, $\mathbf{w}_i^H$ can be derived as

$$\mathbf{w}_i^H = -\lambda^H \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{118}$$

This result reinserted into the constraint in (62) leads to an expression for λ according to

$$\lambda^H = -\begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \left(\underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y} x_i} \right)^{-1} \underline{\mathbf{C}}_{x_i x_i}. \tag{119}$$

Eq. (119) reinserted into (118) leads to the final solution of the optimization problem in the form of

$$\mathbf{w}_{\text{CWL},i}^H = \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \left(\underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y} x_i} \right)^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{120}$$

Appendix B
Derivation of the conditional properties of the CWCU WLMMSE estimator

In appendix B the conditional variance of the estimator proposed in Result 4 will be derived. For simplicity of the formulas we will denote $\mathbf{w}_{\text{CWL},i}^H$ as $\mathbf{w}_i^H$ and $\hat{x}_{\text{CWL},i}$ as $\hat{x}_i$, respectively. We then have

$$\begin{aligned} \text{var}(\hat{x}_i|x_i) &= E\left[|\hat{x}_i - E[\hat{x}_i|x_i]|^2 | x_i \right] \\ &= E\left[|\mathbf{w}_i^H \underline{\mathbf{y}} + b_i - x_i|^2 | x_i \right] \\ &= E\left[|\mathbf{w}_i^H (\underline{\mathbf{y}} - E[\underline{\mathbf{y}}]) - (x_i - E[x_i])|^2 | x_i \right]. \end{aligned}$$

In the last step (60) has been used. Inserting (57) for $E[\underline{\mathbf{y}}]$ leads to

$$\begin{aligned} \text{var}(\hat{x}_i|x_i) = E[&|\mathbf{w}_i^H (\underline{\mathbf{y}} - E[\underline{\mathbf{y}}|x_i]) \\ &+ \mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{y} x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} (\underline{x}_i - E[\underline{x}_i]) \\ &- (x_i - E[x_i])|^2 | x_i]. \end{aligned}$$

Inserting (65) gives

$$\begin{aligned}\mathrm{var}(\hat{x}_i|x_i) =&E[|\mathbf{w}_i^H(\underline{\mathbf{y}} - E[\underline{\mathbf{y}}|x_i])\\ &+ \begin{bmatrix}1 & 0\end{bmatrix}(\underline{x}_i - E[\underline{x}_i]) - (x_i - E[x_i])|^2|x_i]\\ =&E[|\mathbf{w}_i^H(\underline{\mathbf{y}} - E[\underline{\mathbf{y}}|x_i])|^2|x_i]\\ =&\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i}\mathbf{w}_i. \end{aligned} \tag{121}$$

Using

$$\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i} = \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}} \tag{122}$$

(c.f. [12]), and again (65) leads to

$$\begin{aligned}\mathrm{var}(\hat{x}_i|x_i) =& \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}\\ &- \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}\\ =& \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}\\ &- \sigma_{x_i}^2. \end{aligned} \tag{123}$$

Appendix C
Derivation of the CWCU WLMMSE estimator under the jointly Gaussian assumption for real $\mathbf{x}$

In appendix C we solve the optimization problem given in (79) using the Lagrange multiplier method. We start with the Lagrangian cost function which is

$$J' = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}\mathbf{w}_i - \sigma_{x_i}^2 + \lambda\left(\frac{1}{\sigma_{x_i}^2}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\mathbf{w}_i - 1\right). \tag{124}$$

Using Wirtinger's calculus for complex derivatives, the derivation of (124) follows to

$$\frac{\partial J'}{\partial \mathbf{w}_i} = \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^T\mathbf{w}_i^* + (\lambda\frac{1}{\sigma_{x_i}^2}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}})^T. \tag{125}$$

By setting (125) equal to zero, $\mathbf{w}_i^H$ can be derived as

$$\mathbf{w}_i^H = -\lambda\frac{1}{\sigma_{x_i}^2}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}. \tag{126}$$

This result reinserted into the constraint in (79) leads to an expression for λ according to

$$\lambda = -\frac{(\sigma_{x_i}^2)^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}. \tag{127}$$

Eq. (127) reinserted into (126) leads to the final solution of the optimization problem in the form of

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}. \tag{128}$$

Compounding $\mathbf{w}_{\mathrm{CWL},i}^H$ to an estimator matrix immediately leads to Result 5.

Appendix D
Derivation of the conditional properties of the CWCU WLMMSE estimator for real valued parameter vectors

In this appendix the conditional variance of the estimator proposed in Result 5 is investigated. For simplicity of the formulas we will again denote $\mathbf{w}_{\mathrm{CWL},i}^H$ as $\mathbf{w}_i^H$ and $\hat{x}_{\mathrm{CWL},i}$ as $\hat{x}_i$, respectively. The first steps correspond to the ones of appendix B, such that after utilizing (75) and (80) we obtain

$$\mathrm{var}(\hat{x}_i|x_i) = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i}\mathbf{w}_i. \tag{129}$$

To find an expression for $\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i}$ we begin with

$$\mathbf{C}_{\mathbf{y}_\mathbb{R}\mathbf{y}_\mathbb{R}|x_i} = \mathbf{C}_{\mathbf{y}_\mathbb{R}\mathbf{y}_\mathbb{R}} - \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\mathbf{y}_\mathbb{R}}. \tag{130}$$

Multiplying (130) with $\mathbf{T}$ from the left and with $\mathbf{T}^H$ from the right and utilizing (24) yields

$$\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i} = \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\underline{\mathbf{y}}}. \tag{131}$$

Replacing $\mathbf{C}_{\underline{\mathbf{y}}x_i}$ with $\underline{\mathbf{C}}_{\mathbf{y}x_i}\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ finally leads to

$$\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i} = \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}. \tag{132}$$

Inserting (132) into (129) results in

$$\begin{aligned}\mathrm{var}(\hat{x}_i|x_i) =&\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i}\mathbf{w}_i\\ =&\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}\mathbf{w}_i - \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\mathbf{w}_i\\ =&\frac{\left(\sigma_{x_i}^2\right)^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}} - \sigma_{x_i}^2, \end{aligned} \tag{133}$$

where in the last step, (80) has been incorporated.

References

[1] S. M. Kay, "Fundamentals of statistical signal processing: estimation theory," Prentice-Hall PTR, 1st edition, April 2010.

[2] L. Fu-Müller, M. Lunglmayr, M. Huemer, "Reducing the Gap between Linear Biased Classical and Linear Bayesian Estimation," *Proc. IEEE Statistical Signal Processing Workshop (SSP 2012)*, pp. 65-68, Ann Arbor, Michigan, USA, August 2012.

[3] Y.C. Eldar "Rethinking biased estimation improving maximum likelihood and the Cramr-Rao bound", Hanover, Mass. :Now Publishers, 2008.

[4] S. M. Kay, "Fundamentals of Statistical Signal Processing: Volume III," Upper Saddle River, NJ: Prentice Hall, 2013.

[5] M. Triki, D.T.M. Slock, "Component-Wise Conditionally Unbiased Bayesian Parameter Estimation: General Concept and Applications to Kalman Filtering and LMMSE Channel Estimation," In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pp. 670-674, Pacific Grove, USA, Nov. 2005.

[6] M. Triki, A. Salah, D.T.M. Slock, "Interference cancellation with Bayesian channel models and application to TDOA/IPDL mobile positioning," *Proc. International Symposium on Signal Processing and its Applications*, pp.299-302, August 2005.

[7] M. Triki, D.T.M. Slock, "Investigation of Some Bias and MSE Issues in Block-Component-Wise Conditionally Unbiased LMMSE," *Proc. Asilomar Conf. Signals, Syst., Comput.*, pp.1420-1424, Pacific Grove, USA, Nov. 2006.

[8] M. Huemer, O. Lang, "On component-wise conditionally unbiased linear Bayesian estimation," in Signals, Systems and Computers, 2014 48th Asilomar Conference on , vol., no., pp.879-885, 2-5 Nov. 2014

[9] O. Lang, M. Huemer, "CWCU LMMSE Estimation under Linear Model Assumptions," In the Lecture Notes in Computer Science (LNCS): Computer Aided Systems Theory - EUROCAST 2015 (15th International Conference, Las Palmas de Gran Canaria, Spain, February 2015, revised selected papers), Vol. 9520, pp. 537-545, December 2015

[10] T. Adali, P.J. Schreier, L.L. Scharf; "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," *IEEE Trans. Signal Process.*, vol. 59, issue 11, pp.5101-5125, 2011.

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

[12] P.J. Schreier, L.L. Scharf; "Statistical signal processing of complex-valued data: the theory of improper and noncircular signals," 2010,Cambridge University Press.

[13] M. Witzke, "Linear and widely linear filtering applied to iterative detection of generalized MIMO signals," *Ann. Telecommun.*, vol. 60, pp. 147168, 2005.

[14] J. Via, D. Ramirez, I. Santamaria, and L. Vielva, "Widely and semi-widely linear processing of quaternion vectors," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. (IEEE ICASSP), Dallas, TX, Mar. 2010, pp. 39463949.

[15] P. Chevalier, J.P. Delmas, and A. Oukaci, "Optimal widely linear MVDR beamforming for noncircular signals," *Proc. Int. Conf. Acoust., Speech, Signal Process*. (IEEE ICASSP), pp. 35733576, 2009.

[16] A. Mirbagheri, N. Plataniotis, and S. Pasupathy, "An enhanced widely linear CDMA receiver with OQPSK modulation," *IEEE Trans. Commun.*, vol. 54, pp. 261272, 2006.

[17] C. Stanciu, J. Benesty, C. Paleologu, T. Gnsler, S. Ciochin, "A widely linear model for stereophonic acoustic echo cancellation," Signal Processing, Volume 93, Issue 2, February 2013, Pages 511-516, ISSN 0165-1684.

[18] D. Xu, C. Gong, S. Cao, X. Xu, Z. Ye, "Robust widely linear beamforming based on spatial spectrum of noncircularity coefficient," Signal Processing, Volume 104, November 2014, Pages 167-173, ISSN 0165-1684.

[19] 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. 53235336, 2007.

[20] C. Jahanchahi, S. Kanna, D. Mandic, "Complex dual channel estimation: Cost effective widely linear adaptive filtering," Signal Processing, Volume 104, November 2014, Pages 33-42, ISSN 0165-1684.

[21] J. Capon, R.J. Greenfield, R.J. Kolker, "Multidimensional maximum-likelihood processing of a large aperture seismic array," in Proceedings of the IEEE , vol.55, no.2, pp.192-211, Feb. 1967.

[22] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841844, 2008.

[23] M. Novey, T. Adali, and A. Roy, "Circularity and Gaussianity detection using the complex generalized Gaussian distribution," *IEEE Signal Proc. Lett.*, vol. 16, no. 1, pp. 993996, Nov. 2009.

[24] J. Eriksson, E. Ollila, and V. Koivunen, "Essential statistics and tools for complex random variables," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 54005408, 2010.

[25] A.T. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 825834, Mar. 2009.

[26] P.J. Schreier and L.L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714725, Mar. 2003.

[27] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 34733482, Dec. 1994.

[28] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 14271433, Mar. 2010.

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

[30] A. Van Den Bos, "The multivariate complex normal distribution-a generalization," *IEEE Trans. Information Theory*, vol.41, no.2, pp.537-539, Mar 1995.

[31] W. Wirtinger, "Zur formalen Theorie der Funktionen von mehr komplexen Veränderliche," *Mathematische Annalen*, vol.97, pp.357-375, 1927.

Mario Huemer Mario Huemer was born in Wels, Austria, in 1970. He received the Dipl.-Ing. degree in mechatronics and the Dr.techn. (Ph.D.) degree from the Johannes Kepler University (JKU) Linz, Austria, in 1996 and 1999, respectively. From 1997 to 2000, he was a research assistant at the Institute for Communications and Information Engineering at JKU Linz, Austria. From 2000 to 2002, he was with Infineon Technologies Austria, research and development center for wireless products. From 2002-2004 he was a lecturer at the University of Applied Sciences of Upper Austria, and from 2004-2007 he was associate professor for electronics engineering at the University of Erlangen-Nuremberg, Germany. In 2007 Mario Huemer moved to Klagenfurt, Austria, to establish the Chair of Embedded Systems and Signal Processing at Klagenfurt University as a full professor. From 2012 to 2013 he served as dean of the Faculty of Technical Sciences. Since September 2013 he is head of the newly founded Institute of Signal Processing at JKU Linz, Austria.

His research interests are adaptive and statistical signal processing, signal processing architectures and implementations, as well as mixed signal processing with applications in communications, radio frequency and baseband integrated circuits, sensor and biomedical signal processing. Within these fields he published more than 180 papers. In 2000 Mario Huemer received the German ITG and the Austrian GIT award for dissertations, and in 2010 the Austrian Kardinal Innitzer award in natural sciences. His review work includes national and European research projects as well as international journals. Since 2009 he is member of the editorial board of the "International Journal of Electronics and Communications (AEUE)".

Mario Huemer is member of the IEEE Signal Processing Society, the IEEE Circuits and Systems Society, and the IEEE Communications Society . He is also member of the German Society of Information Technology (ITG), and the Austrian Electrotechnical Association (OVE).

Oliver Lang Oliver Lang was born in Schärding, Austria in 1987. He studied Electrical Engineering at the Vienna University of Technology and received his Bachelor degree in 2011. In the next two years Oliver Lang studied Microelectronics at the Vienna University of Technology and finished it with excellent success. The topic of his master thesis was the development and analysis of models for a scanning microwave microscope. Since February 2014, he is a member of the Institute of Signal Processing at the Johannes Kepler University in Linz, Austria.

Christian Hofbauer Christian Hofbauer studied Hardware/Software Systems Engineering at the University of Applied Sciences of Upper Austria from 2002 until 2006. He earned his DI (FH) degree, doing his thesis "High level refinement and implementation cost estimation for Matlab/Simulink models, targeting handheld Software Defined Radio architectures" at the research institute IMEC in Belgium. In 2007, Christian joined the Embedded Systems and Signal Processing Group at the Alpen-Adria-Universität Klagenfurt. From 2013 until 2015, he has been a member of the Institute of Signal Processing at the Johannes Kepler University Linz, focusing his research on the investigation of the UW-OFDM concept. Since September 2015, he is working for the Linz Center of Mechatronics GmbH at the position of a senior researcher.