Signal Processing 133 (2017) 227–239

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Component-wise conditionally unbiased widely linear MMSE estimation

Mario Huemer[a], Oliver Lang[a,*], Christian Hofbauer[b]

[a] Institute of Signal Processing, Johannes Kepler University, 4040 Linz, Austria
[b] Linz Center of Mechatronics GmbH, 4040 Linz, Austria

ARTICLE INFO

Keywords:
Bayesian estimation
Best linear unbiased estimator
Linear minimum mean square error
Widely linear estimation
Component-wise conditionally unbiased
CWCU

ABSTRACT

Biased estimators can outperform unbiased ones in terms of the mean square error (MSE). The best linear unbiased estimator (BLUE) fulfills the so called global conditional unbiased constraint when treated in the Bayesian framework. Recently, the component-wise conditionally unbiased linear minimum mean square error (CWCU LMMSE) estimator has been introduced. This estimator preserves a quite strong (namely the CWCU) unbiased condition which in effect sufficiently represents the intuitive view of unbiasedness. Generally, it is global conditionally biased and outperforms the BLUE in a Bayesian MSE sense. In this work 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 CWCU WLMMSE and WLMMSE estimators. The properties of the CWCU WLMMSE estimator are deduced analytically, and compared by simulation to global conditionally unbiased as well as WLMMSE counterparts with the help of a parameter estimation example and a data estimation/channel equalization application.

1. 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

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

where $p(\mathbf{y}; \mathbf{x})$ is the probability density function (PDF) of vector $\mathbf{y}$ parameterized 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

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

where the integration is performed over the joint PDF of $\mathbf{x}$ and $\mathbf{y}$. Compared to the classical unbiased constraint in (1), (2) is a much softer requirement, which will be particularly discussed in Section 6. 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.

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

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 [5–10], 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 $i = 1, 2, \ldots, n$). The CWCU constraints are less stringent than the global conditional unbiased 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

* Corresponding author.
E-mail address: oliver.lang@jku.at (O. Lang).

http://dx.doi.org/10.1016/j.sigpro.2016.10.018
Received 7 January 2016; Received in revised form 20 October 2016; Accepted 22 October 2016
Available online 26 October 2016
0165-1684/

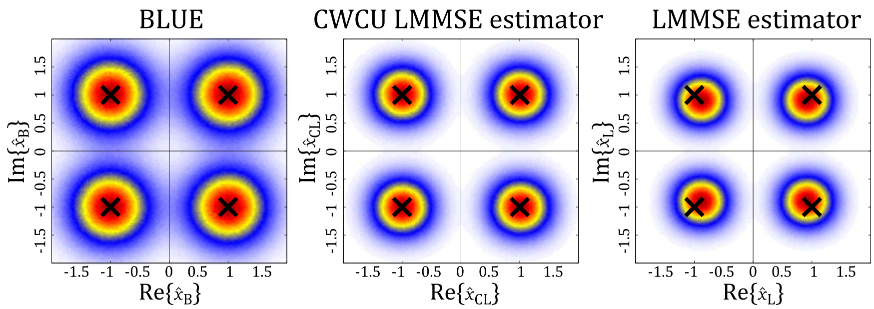

Fig. 1. Visualization of the relative frequencies of the BLUE, the CWCU LMMSE estimator, and the LMMSE estimator, respectively. The black crosses mark the ideal 4-QAM constellation points.

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 unbiased property, which is visualized for a particular example in Fig. 1 (taken from [10]). 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)–(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 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. The LMMSE estimator is conditionally biased towards the prior mean which is 0. The CWCU constraints prevent this bias introduced by the LMMSE estimator, while still allowing the incorporation of prior knowledge about the data in this example that reduces the BMSE compared to the BLUE. Hence, Fig. 1 nicely demonstrates the effects of the CWCU constraints as a trade-off between classical and Bayesian unbiasedness. For details on that example we refer the reader to [10]. Although the BMSEs of the CWCU LMMSE estimator and the LMMSE estimator differ, is has been shown in [8] that the corresponding log-likelihood ratios (LLRs) and consequently the bit error ratios (BERs) coincide for this digital communication example. Possible applications of CWCU estimators include scenarios where a conditional bias as for the LMMSE estimator in Fig. 1 is disadvantageous. Another possible scenario where CWCU estimators may be employed is where they allow for simplifying follow-up processing steps. Such an example is discussed in Section 5.2 where the CWCU WLMMSE estimator allows for simplifying the LLR evaluation compared to the WLMMSE estimator. Examples and applications of the CWCU LMMSE estimator can also be found in [5–7].

The theory of the CWCU LMMSE estimator under linear model assumptions has been discussed in [9,10]. 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 2 and [11]. 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 [11], widely linear estimators are preferable over linear estimators [12]. We anticipate some results to be derived in this paper: The CWCU widely linear MMSE (CWCU WLMMSE) estimator always exists under linear model assumptions, and in the worst case it coincides with the best widely linear unbiased estimator (BWLUE). An example where the CWCU WLMMSE estimator and the BWLUE coincide is when there is only one parameter to be estimated. In that case, the CWCU constraints in (4) correspond to (3). However, in a number of practically interesting situations, the CWCU WLMMSE estimator is able to outperform the BWLUE. For the LMMSE estimator and the 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. The situation is different for CWCU estimators, as will be seen is this work. We investigate model assumptions that allow to find a linear or widely linear CWCU estimator that is able to outperform the BLUE or the BWLUE, respectively. In particular, we will derive the CWCU WLMMSE estimator under the following prerequisites, namely

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

We also address the analytical relationship between CWCU WLMMSE and WLMMSE estimators, which is not as straight forward as the relationship between CWCU LMMSE and LMMSE estimators regarded in [10].

The rest of the paper is organized as follows: In Section 2 we recapitulate the mathematical preliminaries required to derive the linear and particularly the widely linear estimators in this work. In Section 3 we extend linear CWCU estimation by a certain case not handled so far in our former papers. Then we turn to widely linear estimation in Section 4 discussing the derivations of the CWCU WLMMSE estimator for different prerequisites. Section 5 contains a parameter estimation 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. Furthermore, it includes a channel equalization / data estimation example where the CWCU WLMMSE estimator offers certain benefits compared to the WLMMSE estimator. Finally, Section

6 compares all regarded estimators from an optimization point of view.

Notation:

Lower-case bold face variables ($\mathbf{a}, \mathbf{b}, \ldots$) indicate vectors, and upper-case bold face variables ($\mathbf{A}, \mathbf{B}, \ldots$) 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 context we simply write $\mathbf{I}$ and $\mathbf{0}$, respectively. The index $_\mathrm{R}$ of a vector or matrix denotes its real part and the index $_\mathrm{I}$ denotes its imaginary part, e.g., $\mathbf{x}_\mathrm{R} = \mathrm{Re}\{\mathbf{x}\}$ and $\mathbf{x}_\mathrm{I} = \mathrm{Im}\{\mathbf{x}\}$. $E[\cdot]$ 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.

2. 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 [11], where an excellent introduction to improper data and widely linear processing can be found.

2.1. Linear and widely linear transformations

The *complex augmented vector* $\underline{\mathbf{x}}$ of a vector $\mathbf{x} = \mathbf{x}_\mathrm{R} + j\mathbf{x}_\mathrm{I}$ is constructed by stacking $\mathbf{x}$ on top of its complex conjugate $\mathbf{x}^*$, i.e.

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

Augmented vectors are always underlined.

In this work we repeatedly consider widely linear transformations of the form

$$\mathbf{y} = \mathbf{H}_1\mathbf{x} + \mathbf{H}_2\mathbf{x}^*, \tag{7}$$

whose augmented version can easily found to be

$$\underline{\mathbf{y}} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^* & \mathbf{H}_1^* \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix} = \underline{\mathbf{H}}\underline{\mathbf{x}}. \tag{8}$$

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

$$\underline{\mathbf{y}} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_1^* \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix} = \underline{\mathbf{H}}\underline{\mathbf{x}} \tag{9}$$

is a subset of the set of widely linear transformations.

2.2. 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 estimators. 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

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{F}\mathbf{y}^* + \mathbf{b}. \tag{10}$$

In general widely linear estimators are superior to their linear counterparts as soon as the measurements $\mathbf{y}$ turn improper, see [13–24] for some possible applications of widely linear estimators. In the Sections on CWCU WLMMSE estimation we introduce

$$\mathbf{W} = [\mathbf{E} \quad \mathbf{F}] \tag{11}$$

and write (10) usually in the form

$$\hat{\mathbf{x}} = \mathbf{W}\underline{\mathbf{y}} + \mathbf{b}. \tag{12}$$

Another way to express the estimator is its augmented version

$$\underline{\hat{\mathbf{x}}} = \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{F}^* & \mathbf{E}^* \end{bmatrix}\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} + \underline{\mathbf{b}} = \underline{\mathbf{E}}\underline{\mathbf{y}} + \underline{\mathbf{b}}. \tag{13}$$

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 $i = 1, 2, \ldots, n$ and fulfilling the Bayesian unbiased constraint in (2) is given by

$$\hat{\mathbf{x}} = E_\mathbf{x}[\mathbf{x}] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_\mathbf{y}[\mathbf{y}]). \tag{14}$$

It's widely linear counterpart, the WLMMSE estimator, is most compactly written in its augmented form [11,12]

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

Many technical problems are described by the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{16}$$

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 (16) is

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

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

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^* \end{bmatrix}. \tag{18}$$

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 [25]

$$\hat{\mathbf{x}} = (\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H})^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}. \tag{19}$$

We notice, that we require $m \geq n$ and $\mathrm{rank}(\mathbf{H}) = \mathbf{n}$ to ensure the invertibility of $(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H})$. The widely linear counterpart, the BWLUE, can be identified to be [11]

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

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. [11]).

2.3. Statistics of complex-valued random vectors

In order to characterize the second-order statistical properties of $\mathbf{x}$ we start by considering the augmented covariance matrix

$$\underline{\mathbf{C}}_{\mathbf{xx}} = E[(\underline{\mathbf{x}} - E[\underline{\mathbf{x}}])(\underline{\mathbf{x}} - E[\underline{\mathbf{x}}])^H] \tag{21}$$

$$= \begin{bmatrix} \mathbf{C}_{\mathbf{xx}} & \widetilde{\mathbf{C}}_{\mathbf{xx}} \\ \widetilde{\mathbf{C}}_{\mathbf{xx}}^* & \mathbf{C}_{\mathbf{xx}}^* \end{bmatrix} = \underline{\mathbf{C}}_{\mathbf{xx}}^H \in \mathbb{C}^{2n\times 2n}, \tag{22}$$

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 $\widetilde{\mathbf{C}}_{\mathbf{xx}} = E_\mathbf{x}[(\mathbf{x} - E_\mathbf{x}[\mathbf{x}])(\mathbf{x} - E_\mathbf{x}[\mathbf{x}])^T]$ as the complementary covariance matrix. The relationships between $\mathbf{C}_{\mathbf{xx}}$ and $\widetilde{\mathbf{C}}_{\mathbf{xx}}$ with the covariance matrices $\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}}$, $\mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}}$ and the cross covariance matrix $\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}$ are given by

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} + \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}} + j(\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}^T - \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}) = \mathbf{C}_{\mathbf{xx}}^H \tag{23}$$

and

$$\widetilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} - \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}} + j(\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}^T + \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}) = \widetilde{\mathbf{C}}_{\mathbf{xx}}^T, \tag{24}$$

respectively. $\widetilde{\mathbf{C}}_{\mathbf{xx}}$ is sometimes also referred to as pseudo-covariance matrix or conjugate covariance matrix. If $\widetilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{0}$, then the vector $\mathbf{x}$ is called *proper*, otherwise *improper* [26–31]. Consequently, a scalar random variable is proper if the real and imaginary parts are uncorrelated and have equal variances.

It is easy to see that propriety is preserved by strictly linear transformations, which are represented by block-diagonal augmented matrices.

2.4. Gaussian random vectors

To simplify notation we regard zero mean vectors in this subsection. The PDF of complex $\mathbf{x}$ can be written as

$$p(\mathbf{x}) = \frac{1}{\pi^n \sqrt{\det \underline{\mathbf{C}}_{\mathbf{xx}}}} \exp\left\{-\frac{1}{2}\underline{\mathbf{x}}^H \underline{\mathbf{C}}_{\mathbf{xx}}^{-1} \underline{\mathbf{x}}\right\}, \tag{25}$$

cf. [33,34]. This PDF depends algebraically on $\underline{\mathbf{x}}$, i.e., $\mathbf{x}$ and $\mathbf{x}^*$, and can be used for proper and improper $\mathbf{x}$. In this work we call a complex vector $\mathbf{x}$ following this distribution *generalized complex Gaussian*. The simplification that occurs when $\widetilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{0}$ is obvious and leads to the PDF of a complex proper Gaussian random vector $\mathbf{x}$:

$$p(\mathbf{x}) = \frac{1}{\pi^n \det \mathbf{C}_{\mathbf{xx}}} \exp\{-\mathbf{x}^H \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{x}\}. \tag{26}$$

For a scalar Gaussian random variable properness means that the equipotential lines of its PDF plotted in the complex plane are circles. If the equipotential lines are elliptical, then the scalar Gaussian random variable is improper.

3. 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 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. The findings will be summarized in Result 1 (a) below. In addition in Result 1 (b)-(c) we will recapitulate the results from [10], where the CWCU LMMSE estimator has been derived for different linear model assumptions. As such Result 1 sums up the most important insights on model assumptions that allow to find a CWCU LMMSE estimator that is able to generally outperform the BLUE. We note that the BLUE always also fulfills the CWCU constraints (4). Consequently, for model assumptions that allow to find the BLUE, but that are not covered by Result 1, we concurrently are able to find a CWCU estimator (which not necessarily has to be the CWCU LMMSE estimator).

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

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{b}, \qquad \mathbf{E} \in \mathbb{C}^{n \times m}, \mathbf{b} \in \mathbb{C}^n. \tag{27}$$

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

$$\hat{x}_i = \mathbf{e}_i^H \mathbf{y} + b_i, \tag{28}$$

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

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] + b_i. \tag{29}$$

A closer inspection of (29) 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 $\mathbf{y}$. 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

$$E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] = E_{\mathbf{y}}[\mathbf{y}] + (\sigma_{x_i}^2)^{-1} \mathbf{C}_{\mathbf{y}x_i}(x_i - E_{x_i}[x_i]), \tag{30}$$

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 $\sigma_{x_i}^2$ is the variance of x_i. $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma_{x_i}^2 \tag{31}$$

$$E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}] = b_i. \tag{32}$$

Inserting (28), (31) and (32) in the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ immediately leads to the constrained optimization problem

$$\mathbf{e}_{\mathrm{CL},i} = \underset{\mathbf{e}_i}{\operatorname{argmin}}(\mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2) \ \text{ s. t. } \ \mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma_{x_i}^2, \tag{33}$$

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

$$\mathbf{e}_{\mathrm{CL},i}^H = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}} \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1}. \tag{34}$$

Using $\mathbf{E}_{\mathrm{CL}} = [\mathbf{e}_{\mathrm{CL},1}, \mathbf{e}_{\mathrm{CL},2}, \ldots, \mathbf{e}_{\mathrm{CL},n}]^H$ together with (32) and (34) immediately leads us to the estimator summarized in case (a) of

Result 1. *If* $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$ *are*

(a) ***jointly complex proper Gaussian****, or*
(b) *connected via the linear model in (16) and* $\mathbf{x}$ *is* ***complex proper Gaussian*** *with PDF* $\mathcal{CN}(E_{\mathbf{x}}[\mathbf{x}], \mathbf{C}_{\mathbf{xx}})$ *(the PDF of* $\mathbf{n}$ *is otherwise arbitrary), or*
(c) *connected via the linear model in (16) and* $\mathbf{x}$ *has mean* $E_{\mathbf{x}}[\mathbf{x}]$, ***mutually independent elements*** *and covariance matrix* $\mathbf{C}_{\mathbf{xx}} = \operatorname{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_n}^2\}$ *(the joint PDF of* $\mathbf{x}$ *and* $\mathbf{n}$ *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* $i = 1, 2, \ldots, n$ *is given by*

$$\hat{\mathbf{x}}_{\mathrm{CL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CL}}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]), \tag{35}$$

with

$$\mathbf{E}_{\mathrm{CL}} = \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}, \tag{36}$$

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

$$[\mathbf{D}]_{i,i} = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}}. \tag{37}$$

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*

$$\mathbf{C}_{\mathbf{ee},\mathrm{CL}} = \mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CL}}} = \mathbf{C}_{\mathbf{xx}} - \mathbf{A}\mathbf{D} - \mathbf{D}\mathbf{A} + \mathbf{D}\mathbf{A}\mathbf{D}, \tag{38}$$

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*

$$\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i) = E_{\hat{x}_{\mathrm{CL},i}|x_i}\left[\left|\hat{x}_{\mathrm{CL},i} - E_{\hat{x}_{\mathrm{CL},i}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i]\right|^2 |x_i\right] = \mathbf{e}_{\mathrm{CL},i}^H \mathbf{C}_{\mathbf{yy}|x_i} \mathbf{e}_{\mathrm{CL},i} \tag{39}$$

$$= \frac{(\sigma_{x_i}^2)^2}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}} - \sigma_{x_i}^2 \tag{40}$$

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

Case (b) and (c) are derived in [10]. The statements on $\mathbf{e}$, $\mathbf{C}_{\mathbf{ee},\mathrm{CL}}$ and $\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i)$ can simply be proved by inserting in their corresponding definitions, respectively. The conditional variance and the conditional MSE coincide since the conditional bias is zero. Furthermore, the Bayesian MSE and the conditional MSE coincide since the conditional MSE is independent of the parameter value x_i.

Remarks:

- We'd like to emphasize that although Result 1 (a)-(c) result in the same expression for the CWCU LMMSE estimator, the prerequisites in (a), (b), and (c) clearly differ.

- If $\mathbf{y}$ and $\mathbf{x}$ are connected via the linear model in (16) and if $\mathbf{x}$ and $\mathbf{n}$ are both Gaussian, then they are jointly Gaussian. Furthermore, since $[\mathbf{x}^T, \mathbf{y}^T]^T$ is a linear transformation of $[\mathbf{x}^T, \mathbf{n}^T]^T$, $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian, too. Under these prerequisites it is clear that Result 1 (a) can be applied. However, the main additional message of Result 1 (b) is, that the jointly Gaussian assumption for $\mathbf{x}$ and $\mathbf{n}$ can significantly be relaxed [10]. In fact, the PDF of the noise $\mathbf{n}$ can be arbitrary in case (b).
- The main message of Result 1 (c) is, that for mutually independent elements of the parameter vector also the Gaussian assumption of $\mathbf{x}$ can be abandoned.
- We note, that for the linear model in (16) the covariance matrices required in (36) and (37) become

$$\mathbf{C}_{x_i\mathbf{y}} = \mathbf{C}_{x_i\mathbf{x}}\mathbf{H}^H \tag{41}$$

$$\mathbf{C}_{\mathbf{y}x_i} = \mathbf{H}\mathbf{C}_{\mathbf{x}x_i} \tag{42}$$

$$\mathbf{C}_{\mathbf{yy}} = \mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}} \tag{43}$$

$$\mathbf{C}_{\mathbf{xy}} = \mathbf{C}_{\mathbf{xx}}\mathbf{H}^H. \tag{44}$$

- From (36) 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

$$\mathbf{D} = \mathrm{diag}\{\mathbf{C}_{\mathbf{xx}}\}(\mathrm{diag}\{\mathbf{A}\})^{-1}. \tag{45}$$

- The CWCU LMMSE estimator will in general not commute over linear transformations, an exception is discussed in [9].

4. CWCU WLMMSE estimation

We again assume, that a parameter vector $\mathbf{x}$ is to be estimated based on a measurement vector $\mathbf{y} \in \mathbb{C}^m$. 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). As in the previous section we investigate the estimator under different model assumptions, and we distinguish real and complex valued parameters for reasons that become clear soon. Similar as in Result 1 above we compactly summarize the findings for different model assumptions in Result 2 and 3, respectively.

4.1. Complex parameter vectors

We begin with the assumption that complex $\mathbf{y}$ and complex $\mathbf{x}$ are generalized jointly Gaussian. Further, we assume the widely linear estimator for x_i to be of the form

$$\hat{x}_i = \mathbf{e}_i^H\mathbf{y} + \mathbf{f}_i^H\mathbf{y}^* + b_i, \quad \text{for } i = 1, 2, \ldots, n. \tag{46}$$

Eq. (46) can also be written as

$$\hat{x}_i = \mathbf{w}_i^H\underline{\mathbf{y}} + b_i, \quad \text{for } i = 1, 2, \ldots, n, \tag{47}$$

where we used

$$\mathbf{w}_i^H = [\mathbf{e}_i^H \quad \mathbf{f}_i^H]. \tag{48}$$

The conditional mean of the estimator in (47) follows to

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

Because of the generalized jointly Gaussian assumption on $\mathbf{y}$ and $\mathbf{x}$, $E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ is linear in $\underline{\mathbf{x}}_i = [x_i \quad x_i^*]^T$, namely

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]). \tag{50}$$

This leads to

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H(E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])) + b_i. \tag{51}$$

By setting (51) equal to $x_i = [1 \quad 0]\underline{\mathbf{x}}_i$ we find that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1} = [1 \quad 0] \tag{52}$$

$$E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] = b_i. \tag{53}$$

These are the two conditions the widely linear estimator in (47) 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

$$\begin{aligned} J &= E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2] = E_{\mathbf{y},\mathbf{x}}[|\mathbf{w}_i^H\underline{\mathbf{y}} + b_i - x_i|^2] = E_{\mathbf{y},\mathbf{x}}[|\mathbf{w}_i^H(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - (x_i - E_{x_i}[x_i])|^2] \\ &= E_{\mathbf{y},\mathbf{x}}[|\mathbf{w}_i^H(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - [1 \quad 0](\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])|^2] = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix} \\ &\quad -[1 \quad 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\mathbf{w}_i + \underbrace{[1 \quad 0]\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}}_{\sigma_{x_i}^2}. \end{aligned} \tag{54}$$

This result can be simplified by using (52), leading to the final optimization problem

$$\mathbf{w}_{\mathrm{CWL},i} = \underset{\mathbf{w}_i}{\mathrm{argmin}}(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2) \quad \text{s. t.} \quad \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1} = [1 \quad 0], \tag{55}$$

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 lead to the estimator summarized in case (a) of

Result 2. *If $\mathbf{x} \in \mathbb{C}^n$ is a* ***complex valued parameter vector*** *and*

(a) *$\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^m$ are* ***generalized jointly Gaussian*** *, or*
(b) *$\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^m$ are connected via the linear model in (16) and $\mathbf{x}$ is* ***generalized complex Gaussian*** *with mean vector $E_{\mathbf{x}}[\mathbf{x}]$ and augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{xx}}$ (the PDF of $\mathbf{n}$ is otherwise arbitrary), or*
(c) *$\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^m$ are connected via the linear model in (16) and $\mathbf{x}$ has mean $E_{\mathbf{x}}[\mathbf{x}]$ and* ***mutually independent elements*** *such that $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_n}^2\}$ and $\tilde{\mathbf{C}}_{\mathbf{xx}} = \mathrm{diag}\{\tilde{\sigma}_{x_1}^2, \tilde{\sigma}_{x_2}^2, \ldots, \tilde{\sigma}_{x_n}^2\}$ (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 $i = 1, 2, \ldots, n$ is

$$\hat{\mathbf{x}}_{\mathrm{CWL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{W}_{\mathrm{CWL}}(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]), \tag{56}$$

with

$$\mathbf{W}_{\mathrm{CWL}} = [\mathbf{w}_{\mathrm{CWL},1} \quad \mathbf{w}_{\mathrm{CWL},2} \quad \cdots \quad \mathbf{w}_{\mathrm{CWL},n}]^H, \tag{57}$$

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

$$\mathbf{w}_{\mathrm{CWL},i}^H = [1 \quad 0]\underline{\mathbf{C}}_{x_ix_i}(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{58}$$

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

$$\begin{aligned} \mathbf{C}_{\mathbf{ee},\mathrm{CWL}} &= \mathbf{C}_{\mathbf{xx}} - \mathbf{W}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yx}}\begin{bmatrix}\mathbf{I}^{n\times n}\\ \mathbf{0}^{n\times n}\end{bmatrix} - [\mathbf{I}^{n\times n} \quad \mathbf{0}^{n\times n}]\underline{\mathbf{C}}_{\mathbf{xy}}\mathbf{W}_{\mathrm{CWL}}^H \\ &\quad + \mathbf{W}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{W}_{\mathrm{CWL}}^H. \end{aligned} \tag{59}$$

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

$$\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) = E[|\hat{x}_{\mathrm{CWL},i} - E[\hat{x}_{\mathrm{CWL},i}|x_i]|^2|x_i] = \mathbf{w}_{\mathrm{CWL},i}^H\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\mathbf{w}_{\mathrm{CWL},i} \tag{60}$$

$$=[1\ \ 0]\underline{\mathbf{C}}_{x_i x_i}(\underline{\mathbf{C}}_{x_i \mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}\underline{\mathbf{C}}_{x_i x_i}\begin{bmatrix}1\\0\end{bmatrix}-\sigma_{x_i}^2. \tag{61}$$

The statements on $\mathbf{e}$ and $\mathbf{C}_{\mathbf{ee},\mathrm{CWL}}$ can simply be proved by inserting in their corresponding definitions, respectively. The derivation of the conditional variance can be found in Appendix B. Due to similar arguments as for the CWCU LMMSE estimator in Section 3, the conditional variance, the conditional MSE, and the Bayesian MSE coincide.

We now consider case (b) and (c) of Result 2. Let $\mathbf{x}$ and $\mathbf{y}$ be connected via the linear model (16) (or its augmented version (17)). In the following it will be seen that some of the prerequisites of Result 2 (a) can be relaxed when incorporating details of the data model into the derivation of the estimator.

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. Under these prerequisites it is clear that Result 2 (a) can be applied. We note, that for the linear model the augmented covariance matrices required in (58) and (59) become

$$\underline{\mathbf{C}}_{x_i\mathbf{y}} = \underline{\mathbf{C}}_{x_i\mathbf{x}}\underline{\mathbf{H}}^H \tag{62}$$

$$\underline{\mathbf{C}}_{\mathbf{y}x_i} = \underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{x}x_i} \tag{63}$$

$$\underline{\mathbf{C}}_{\mathbf{yy}} = \underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H + \underline{\mathbf{C}}_{\mathbf{nn}} \tag{64}$$

$$\underline{\mathbf{C}}_{\mathbf{xy}} = \underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H. \tag{65}$$

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. Let $\mathbf{h}_i \in \mathbb{C}^m$ be the i^{th} column of $\mathbf{H}$, $\overline{\mathbf{H}}_i \in \mathbb{C}^{m\times(n-1)}$ the matrix resulting from $\mathbf{H}$ by deleting $\mathbf{h}_i$, x_i be the i^{th} element of $\mathbf{x}$, and $\overline{\mathbf{x}}_i \in \mathbb{C}^{(n-1)}$ the vector resulting from $\mathbf{x}$ after deleting x_i. Then we can rewrite (16) as

$$\mathbf{y} = \mathbf{h}_i x_i + \overline{\mathbf{H}}_i\overline{\mathbf{x}}_i + \mathbf{n}. \tag{66}$$

With the notation

$$\underline{\mathbf{H}}_i = \begin{bmatrix}\mathbf{h}_i & \mathbf{0}\\ \mathbf{0} & \mathbf{h}_i^*\end{bmatrix} \in \mathbb{C}^{2m\times 2}, \quad \underline{\overline{\mathbf{H}}}_i = \begin{bmatrix}\overline{\mathbf{H}}_i & \mathbf{0}\\ \mathbf{0} & \overline{\mathbf{H}}_i^*\end{bmatrix} \in \mathbb{C}^{2m\times(2n-2)} \tag{67}$$

the augmented form of (66) follows to

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\overline{\mathbf{H}}}_i\underline{\overline{\mathbf{x}}}_i + \underline{\mathbf{n}}. \tag{68}$$

Incorporating (68) into the conditional mean of the estimator in (47) yields

$$\begin{aligned}E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] &= E_{\mathbf{y}|x_i}[\mathbf{w}_i^H\underline{\mathbf{y}} + b_i|x_i] = E_{\mathbf{n},\overline{\mathbf{x}}_i|x_i}[\mathbf{w}_i^H(\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\overline{\mathbf{H}}}_i\underline{\overline{\mathbf{x}}}_i + \underline{\mathbf{n}}) + b_i|x_i]\\ &= \mathbf{w}_i^H(\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\overline{\mathbf{H}}}_i E_{\overline{\mathbf{x}}_i|x_i}[\underline{\overline{\mathbf{x}}}_i|x_i]) + b_i.\end{aligned} \tag{69}$$

From (69) 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 $i = 1, 2, \ldots, n$) if $\mathbf{x}$ is generalized complex Gaussian (the reasoning follows immediately). This will lead us to Result 2 (b).
2. (4) can be fulfilled for all possible x_i (and all $i = 1, 2, \ldots, n$) if $E_{\overline{\mathbf{x}}_i|x_i}[\underline{\overline{\mathbf{x}}}_i|x_i] = E_{\overline{\mathbf{x}}_i}[\underline{\overline{\mathbf{x}}}_i]$ for all possible x_i (and all $i = 1, 2, \ldots, n$), which is true if the elements x_i of $\mathbf{x}$ are mutually independent. This will lead us to Result 2 (c).
3. (4) is fulfilled for all possible x_i (and all $i = 1, 2, \ldots, n$) if $\mathbf{w}_i^H\underline{\mathbf{H}}_i = [1\ \ 0]$ and $\mathbf{w}_i^H\underline{\overline{\mathbf{H}}}_i = \mathbf{0}^T$ for $i = 1, 2, \ldots, n$, and if we set $b_i = 0$. These constraints and settings correspond to the ones of the BWLUE [11]. Consequently, the BWLUE is a CWCU estimator.

We start with the first option, 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

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H(\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\overline{\mathbf{H}}}_i(E_{\overline{\mathbf{x}}_i}[\underline{\overline{\mathbf{x}}}_i] + \underline{\mathbf{C}}_{\overline{\mathbf{x}}_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]))) + b_i. \tag{70}$$

By setting (70) equal to $x_i = [1\ \ 0]\underline{\mathbf{x}}_i$ one can see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H\underline{\mathbf{H}}_i + \mathbf{w}_i^H\underline{\overline{\mathbf{H}}}_i\underline{\mathbf{C}}_{\overline{\mathbf{x}}_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1} = [1\ \ 0], \tag{71}$$

$$b_i = -\mathbf{w}_i^H\underline{\overline{\mathbf{H}}}_i(E_{\overline{\mathbf{x}}_i}[\underline{\overline{\mathbf{x}}}_i] - \underline{\mathbf{C}}_{\overline{\mathbf{x}}_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}E_{x_i}[\underline{\mathbf{x}}_i]). \tag{72}$$

After some algebraic manipulations (71) and (72) can compactly be written as

$$\mathbf{w}_i^H\underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{x}x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1} = [1\ \ 0], \tag{73}$$

$$b_i = E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. \tag{74}$$

Eq. (73) could also have been derived from (52) 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

$$\mathbf{w}_{\mathrm{CWL},i} = \underset{\mathbf{w}_i}{\operatorname{argmin}}(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2)\quad \text{s. t.}\quad \mathbf{w}_i^H\underline{\mathbf{H}}\underline{\mathbf{C}}_{\mathbf{x}x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1} = [1\ \ 0]. \tag{75}$$

The solution to this constrained optimization problem can be found using the Lagrange multiplier method, which finally leads to the statements of Result 2 (b).

For mutually independent parameters (case (c) of Result 2) it is possible to further relax the prerequisites on $\mathbf{x}$. In this case (69) becomes

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \mathbf{w}_i^H\underline{\overline{\mathbf{H}}}_i E_{\overline{\mathbf{x}}_i}[\underline{\overline{\mathbf{x}}}_i] + b_i, \tag{76}$$

since $E_{\overline{\mathbf{x}}_i|x_i}[\underline{\overline{\mathbf{x}}}_i|x_i]$ is no longer dependent on x_i. By setting (76) equal to $x_i = [1\ \ 0]\underline{\mathbf{x}}_i$ we see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{w}_i^H\underline{\mathbf{H}}_i = [1\ \ 0], \tag{77}$$

$$b_i = -\mathbf{w}_i^H\underline{\overline{\mathbf{H}}}_i E_{\overline{\mathbf{x}}_i}[\underline{\overline{\mathbf{x}}}_i] = E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. \tag{78}$$

Eq. (77) could also have been derived from (71) 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

$$\mathbf{w}_{\mathrm{CWL},i} = \underset{\mathbf{w}_i}{\operatorname{argmin}}(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2)\quad \text{s. t.}\quad \mathbf{w}_i^H\underline{\mathbf{H}}_i = [1\ \ 0]. \tag{79}$$

The solution to this constrained optimization problem can again be found using the Lagrange multiplier method, which finally leads to the statements of Result 2 (c).

The CWCU WLMMSE estimator matrix $\mathbf{W}_{\mathrm{CWL}}$ from **Result 2** 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

$$\mathbf{W}_{\mathrm{CWL}} = [\mathbf{D}_1\ \ \mathbf{D}_2]\underline{\mathbf{E}}_{\mathrm{WL}}, \tag{80}$$

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

$$[\mathbf{D}_1]_{i,i} = [[1\ \ 0]\underline{\mathbf{C}}_{x_i x_i}(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}]_{1,1}, \tag{81}$$

$$[\mathbf{D}_2]_{i,i} = [[1\ \ 0]\underline{\mathbf{C}}_{x_i x_i}(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i})^{-1}]_{1,2}. \tag{82}$$

The complexity of applying the CWCU WLMMSE estimator and the WLMMSE estimator is the same since both estimators are of the widely linear form. However, deriving the CWCU WLMMSE estimator matrix is slightly more complex than for the WLMMSE estimator, as one can see in (80)–(82). Eq. (81) and (82) require $\underline{\mathbf{C}}_{x_i\mathbf{y}}$ and the inverse of $\underline{\mathbf{C}}_{\mathbf{yy}}$. Both terms can be reused from the calculation of the WLMMSE estimator matrix $\underline{\mathbf{E}}_{\mathrm{WL}} = \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}$. Furthermore, the inverse of $\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}$ is required for $i = 1, 2, \ldots, n$, however, $\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}$ is just a 2×2 matrix.

We finally notice, that in a linear model scenario, that does not fulfill the assumptions of Result 2 (b) or (c) we still can derive the BWLUE which is a CWCU estimator.

4.2. Real parameter vectors

In this subsection 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 for all three cases of Result 2, since

$$\underline{\mathbf{C}}_{x_ix_i} = \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{x_i}^2 \\ \sigma_{x_i}^2 & \sigma_{x_i}^2 \end{bmatrix}. \tag{83}$$

However, we now assume the real composite vector

$$\mathbf{y}_{\mathbb{R}} = \begin{bmatrix} \mathbf{y}_{\mathrm{R}} \\ \mathbf{y}_{\mathrm{I}} \end{bmatrix} \in \mathbb{R}^{2m}, \tag{84}$$

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

$$E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i] = E_{\mathbf{y}_{\mathbb{R}}}[\mathbf{y}_{\mathbb{R}}] + \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \tag{85}$$

By multiplying (85) with the real-to-complex transformation matrix

$$\mathbf{T}_n = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \in \mathbb{C}^{2n\times 2n} \tag{86}$$

from the left we obtain an expression for $E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ according to

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \tag{87}$$

With (87) the conditional mean of the estimator in (47) becomes

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{w}_i^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] + b_i = \mathbf{w}_i^H\left(E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i])\right) + b_i. \tag{88}$$

By setting (88) 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

$$\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\frac{1}{\sigma_{x_i}^2} = 1 \tag{89}$$

$$E_{x_i}[x_i] - \mathbf{w}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] = b_i. \tag{90}$$

Simplifying the BMSE cost function in (54) using the constraint in (89) leads to the optimization problem

$$\mathbf{w}_{\mathrm{CWL},i} = \underset{\mathbf{w}_i}{\operatorname{argmin}}(\mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \sigma_{x_i}^2) \quad \text{s. t.} \quad \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\frac{1}{\sigma_{x_i}^2} = 1. \tag{91}$$

The solution of this optimization problem is derived in Appendix C. The results of Appendix C lead to the estimator summarized in case (a) of

Result 3. *Let* $\mathbf{y} \in \mathbb{C}^m$. *If* $\mathbf{x} \in \mathbb{R}^n$ *is a* ***real valued parameter vector*** *and*

(a) $\mathbf{x}$ *and* $\mathbf{y}_{\mathbb{R}} \in \mathbb{R}^{2m}$ *are* ***jointly Gaussian*** *, or*
(b) $\mathbf{x}$ *and* $\mathbf{y}$ *are connected via the linear model in* (16) *and* $\mathbf{x}$ *is* ***Gaussian*** *with PDF* $\mathcal{N}(E_{\mathbf{x}}[\mathbf{x}], \mathbf{C}_{\mathbf{xx}})$ *(the PDF of* $\mathbf{n}$ *is otherwise arbitrary), or*
(c) $\mathbf{x}$ *and* $\mathbf{y}$ *are connected via the linear model in* (16) *and* $\mathbf{x}$ *has mean* $E_{\mathbf{x}}[\mathbf{x}]$, ***mutually independent elements*** *and covariance matrix* $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_n}^2\}$ *(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* $i = 1, 2, \ldots, n$ *is given by* (56) *where the rows of* $\mathbf{W}_{\mathrm{CWL}}$ *are given by*

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}}[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{92}$$

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*

$$\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\mathbf{w}_i = \frac{(\sigma_{x_i}^2)^2}{[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}} - \sigma_{x_i}^2. \tag{93}$$

Case (b) and (c) can be derived by following similar steps as in Section 4.1. The derivation of the conditional variances can be found in Appendix D. An alternative representation of (92) can be obtained by utilizing $[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}} = \mathbf{C}_{x_i\underline{\mathbf{y}}}$, yielding

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{C}_{\underline{\mathbf{y}}x_i}}\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{94}$$

The CWCU WLMMSE estimator matrix $\mathbf{W}_{\mathrm{CWL}}$ from Result 3 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

$$\mathbf{W}_{\mathrm{CWL}} = \mathbf{D}[\mathbf{I}^{n\times n}\ \ \mathbf{0}^{n\times n}]\underline{\mathbf{E}}_{\mathrm{WL}}, \tag{95}$$

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

$$[\mathbf{D}]_{i,i} = \frac{\sigma_{x_i}^2}{[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}}. \tag{96}$$

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

4.3. PWCU WLMMSE estimation

We come back to the case of a complex parameter vector $\mathbf{x} \in \mathbb{C}^n$. Another way to estimate $\mathbf{x}$ is to rewrite the linear model $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$ according to

$$\mathbf{y} = \underbrace{[\mathbf{H}\ \ j\mathbf{H}]}_{\mathbf{H}'\in\mathbb{C}^{m\times 2n}}\underbrace{\begin{bmatrix} \mathbf{x}_{\mathrm{R}} \\ \mathbf{x}_{\mathrm{I}} \end{bmatrix}}_{\mathbf{x}_{\mathbb{R}}\in\mathbb{R}^{2n}} + \mathbf{n}, \tag{97}$$

and estimate the real and imaginary parts of the parameter vector separately. With (97), 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 $x_i = x_{\mathrm{R},i} + jx_{\mathrm{I},i}$, but it is a CWCU estimator for $x_{\mathrm{R},i}$ and $x_{\mathrm{I},i}$, since we forced $E[\hat{x}_{\mathrm{R},i}|x_{\mathrm{R},i}] = x_{\mathrm{R},i}$ and $E[\hat{x}_{\mathrm{I},i}|x_{\mathrm{I},i}] = x_{\mathrm{I},i}$ for $i = 1, 2, \ldots, n$. This is why this estimator will be denoted as part-wise conditionally unbiased WLMMSE (PWCU WLMMSE) estimator. Generally, this estimator features a lower BMSE compared to its CWCU counterpart, since conditioning separately on the real and on the imaginary parts generally leads to weaker constraints than when conditioning on the complex parameters. However, there exist cases where CWCU and PWCU estimators feature the same BMSE performance.

5. Examples

5.1. DC level and complex exponential in uncorrelated Gaussian noise

In this example we apply the CWCU WLMMSE estimator to a particular signal parameter estimation problem, and compare it in performance to 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 $k = 0, 1, \ldots, 5$, which can easily be brought to the form of a linear model $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$. We assume the noise vector $\mathbf{n}$ to be complex proper Gaussian $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_{nn}})$ with

$$\mathbf{C_{nn}} = \text{diag}\{0.1,\ 0.06,\ 0.3,\ 0.2,\ 0.15,\ 0.1\}. \tag{98}$$

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

$$\mathbf{C}_{\mathbf{x_R x_R}} = \text{diag}\{1,\ 0.6\} \tag{99}$$

$$\mathbf{C}_{\mathbf{x_I x_I}} = k\ \text{diag}\{1,\ 0.6\} \tag{100}$$

$$\mathbf{C}_{\mathbf{x_R x_I}} = \mathbf{0}^{2\times 2}, \tag{101}$$

where the scalar k in $\mathbf{C}_{\mathbf{x_I x_I}}$ 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 1. Then we increase k stepwise, such that the imaginary part of $\mathbf{x}$ becomes more and more significant, and repeat the estimation procedures accordingly. 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.

Table 1
Estimators used for the problem described in Section 5.1.

Estimator	*Section*	*Equation*
BLUE	2.2	(19)
LMMSE	2.2	(14)
CWCU LMMSE	3	Result 1
BWLUE	2.2	(20)
WLMMSE	2.2	(15)
CWCU WLMMSE	4.1	Result 2
CWCU WLMMSE for real parameter vectors	4.2	Result 3
PWCU WLMMSE	4.3	Result 3

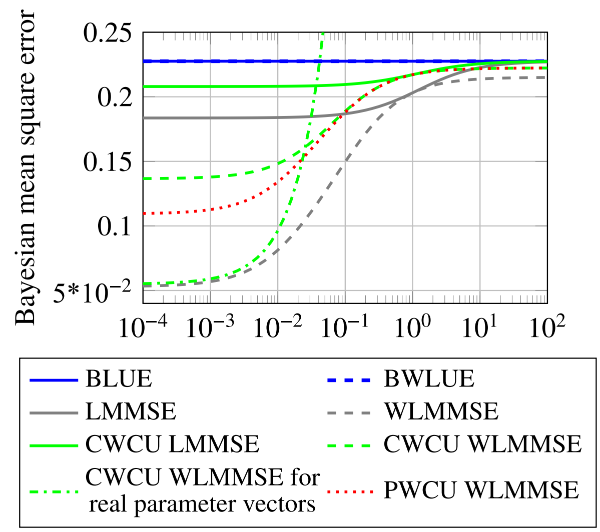

Fig. 2. BMSE values plotted over the scaling factor k which defines the variances of the imaginary parts. The variances of the real parts have been kept constant.

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

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 $\mathbf{C}_{\mathbf{x_I x_I}}$ become less important. What's left is the prior knowledge about $\mathbf{C}_{\mathbf{x_R x_R}}$. 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.

5.2. Estimation of 8-QAM symbols in a unique word OFDM framework

An example where employing the CWCU WLMMSE estimator allows for reducing the computational complexity of a follow-up processing step is presented in this section. In digital communications, data symbols have to be estimated based on the received signal. In this data estimation / channel equalization example we choose 8-QAM data symbols from the alphabet $\mathcal{S} = \{-3 \pm j, -1 \pm j, 1 \pm j, 3 \pm j\}$, which results in improper symbols since the variance of the real part is larger than that of the imaginary part. The following investigations and simulations are carried out within the framework of unique word orthogonal frequency division multiplexing (UW-OFDM) described in [35,36]. Like classical OFDM, UW-OFDM is a block based transmission scheme where at the receive side a data vector $\mathbf{d}$ is estimated based

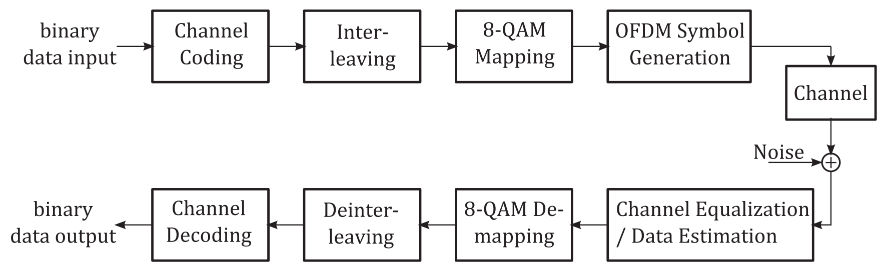

Fig. 3. Block diagram of the investigated UW-OFDM communication system.

on a received block $\tilde{\mathbf{y}}$ of frequency domain samples which are disturbed by a dispersive channel and additive noise. We choose UW-OFDM since the estimator matrices are in general full matrices instead of diagonal matrices as in classical OFDM, such that the problem can be considered a more demanding and general one compared to the data estimation problem in classical OFDM systems. Hence, this framework is well suited for studying general effects of CWCU estimators. The system model for the transmission of one data block is given by

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{G}\mathbf{d} + \tilde{\mathbf{v}}, \tag{102}$$

where $\tilde{\mathbf{H}}$ is the diagonal channel matrix with the frequency response coefficients of the channel on its main diagonal. $\mathbf{G}$ is a so called generator matrix, for details cf. [35,36], $\mathbf{d}$ is a vector of improper 8-QAM symbols and $\tilde{\mathbf{v}}$ is a frequency domain noise vector. A block diagram of the simulation setup is shown in Fig. 3. The first block is implemented as a convolutional encoder with the industry standard rate 1/2, constraint length 7 code with generator polynomials (133, 171) as defined in [37]. The interleaver re-sorts the bits appropriately, which are then mapped onto improper 8-QAM symbols. These symbols are arranged in blocks, each block is converted into an UW-OFDM time domain symbol, and a burst of UW-OFDM symbols is transmitted over the channel. The channel is assumed to be quasi-static, meaning that it stays constant during the transmission of one burst. Furthermore, we assumed perfect channel knowledge at the receiver in these simulations. The widely linear estimators are then applied on each individual received frequency-domain vector $\tilde{\mathbf{y}}$ in order to equalize the channel or rather estimate the data symbols. The 8-QAM demapper determines the LLRs of the corresponding bits and feeds them into the deinterleaver. Finally, a soft decision Viterbi algorithm is applied for decoding.

In our simulation setup the dimensions of the vectors and matrices are as follows: $\mathbf{d} \in \mathbb{C}^{36\times 1}$, $\mathbf{G} \in \mathbb{C}^{52\times 36}$, $\tilde{\mathbf{H}} \in \mathbb{C}^{52\times 52}$, $\tilde{\mathbf{y}} \in \mathbb{C}^{52\times 1}$. The particular generator matrix $\mathbf{G}'$ introduced and described in [35,36] has been used.

For a general estimator, the LLRs of any symbol constellation with equiprobable transmit symbols can be written as [8,38]

$$\Lambda(b_{ki}|\hat{x}_i) = \log\frac{\Pr(b_{ki}=1|\hat{x}_i)}{\Pr(b_{ki}=0|\hat{x}_i)} = \log\frac{\sum_{q\in S(b_{ki}=1)} p(\hat{x}_i|s^{(q)})}{\sum_{q\in S(b_{ki}=0)} p(\hat{x}_i|s^{(q)})}, \tag{103}$$

where $\hat{x}_i$ is the i^{th} estimated symbol, b_{ki} is the k^{th} bit of the i^{th} estimated symbol, $S(b_{ki}=1)$ and $S(b_{ki}=0)$ are the sets of symbol indices corresponding to $b_{ki}=1$ and $b_{ki}=0$, respectively, and $s^{(q)}$ is the q^{th} out of 8 possible 8-QAM symbols. In (103), $p(\hat{x}_i|s^{(q)})$ denotes the conditional PDF of the estimate $\hat{x}_i$ given that the actual symbol was $s^{(q)}$. With (103), the LLRs of any (widely) linear estimator can be evaluated by incorporating $p(\hat{x}_i|s^{(q)})$ for the specific estimator. Such a specific estimator could e.g. be the WLMMSE estimator or the CWCU WLMMSE estimator. Due to central limit theorem arguments (note that the data vector length is 36 in our example) $p(\hat{x}_i|s^{(q)})$ can be well approximated as Gaussian in both cases. If the estimates are improper, the general complex Gaussian density function [11,32–34]

$$p(\hat{x}_i|s^{(q)}) = \frac{1}{\sqrt{\pi^2 \det(\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)}})}}\cdot e^{-\frac{1}{2}(\hat{\underline{\mathbf{x}}}_i - E[\hat{\underline{\mathbf{x}}}_i|s^{(q)}])^H \underline{\mathbf{C}}^{-1}_{\hat{x}_i\hat{x}_i|s^{(q)}}(\hat{\underline{\mathbf{x}}}_i - E[\hat{\underline{\mathbf{x}}}_i|s^{(q)}])} \tag{104}$$

has to be used, otherwise the simpler complex proper Gaussian density

$$p(\hat{x}_i|s^{(q)}) = \frac{1}{\pi\sigma^2_{\hat{x}_i|s^{(q)}}} e^{-\frac{1}{\sigma^2_{\hat{x}_i|s^{(q)}}}|\hat{x}_i - E[\hat{x}_i|s^{(q)}]|^2} \tag{105}$$

can be employed instead, where $\sigma^2_{\hat{x}_i|s^{(q)}}$ denotes the conditional variance of the estimate $\hat{x}_i$ given the transmitted symbol $s^{(q)}$. Note that in contrast to (104), (105) does not require the augmented form, consequently no evaluations of determinants and no matrix inversions are required.

It has been shown in [8], that for the estimated 8-QAM symbols transmitted over an additive white Gaussian noise (AWGN) channel (i.e. $\tilde{\mathbf{H}} = \mathbf{I}$) $p(\hat{x}_i|s^{(q)})$ is proper for the CWCU WLMMSE estimator and improper for the WLMMSE estimator. This result is also suggested by Fig. 4, which is taken from [8]. For the CWCU WLMMSE estimator the estimates are centered around the true constellation points since it fulfills the CWCU constraints. Furthermore, the estimates conditioned on a specific transmit symbol $s^{(q)}$ are properly distributed. In contrast to the CWCU WLMMSE estimates, the WLMMSE estimates conditioned on a specific transmit symbol are neither centered around the true constellation points nor are they properly distributed, cf. Fig. 4b. As a consequence, the CWCU WLMMSE estimator allows for utilizing (105), while the WLMMSE estimator requires (104) to derive the LLRs for further processing. Furthermore, it has been shown in [8] that the LLRs and consequently the BERs of the CWCU WLMMSE estimator and the WLMMSE estimator coincide. Hence, one can conclude that applying the CWCU WLMMSE estimator in this system setup has the advantage of a reduced complexity of the LLR determination compared

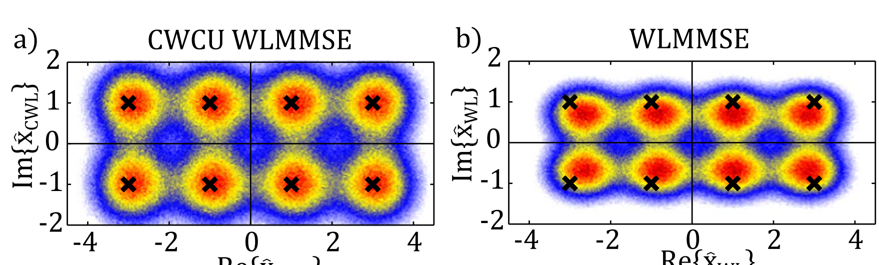

Fig. 4. Relative frequencies of the CWCU WLMMSE estimates in (a), and the WLMMSE estimates in (b). The black crosses mark the original 8-QAM constellation points. Taken from [8].

to the WLMMSE estimator without any loss in the BER. We notice, that (e.g. in WLAN scenarios) the data estimator only has to be derived once per burst, such that the slightly increased complexity of deriving the CWCU WLMMSE estimator matrix is negligible. On the other hand, the LLRs have to be calculated for every single data bit.

In the following we consider multipath channels instead of the AWGN channel. The channel impulse responses (CIRs) are modeled as tapped delay lines, each tap with uniformly distributed phase and Rayleigh distributed magnitude, and with power decaying exponentially as defined in [39]. The model allows the choice of the channel delay spread, for a more detailed description we refer to [39]. 10000 CIR realizations featuring a channel delay spread of $\tau_{\mathrm{RMS}} = 100$ ns have been generated and stored, and the BER simulation results are obtained by averaging over these 10000 realizations.

The data estimation has been performed with the WLMMSE estimator, the CWCU WLMMSE estimator and the BWLUE. (Note, that the BLUE would have the same BER performance as the BWLUE since these estimators cannot utilize the improperness of the data.) It turns out, that the CWCU WLMMSE estimates conditioned on a given transmit symbol $s^{(q)}$ are practically proper again for all channel realizations. The off-diagonal elements of $\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)}}$ are below the main diagonal elements by at least a factor of 10^{-3} in all cases. We therefore again apply (105) for the LLR calculation in case the CWCU WLMMSE estimator is used. The effects on the BER performance in dependency on the mean energy per bit to noise power spectral density ratio E_b/N_0 is visualized in Fig. 5. This figure shows that the loss in performance of the CWCU WLMMSE estimator using the simplified PDF in (105) for LLR calculation is definitely insignificant. Note that in practice usually approximation formulas are used to derive LLRs, in our application this means that (105) instead of (104) can be used as starting point to derive LLR approximations [40–42].

6. 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 variances of the estimator's elements subject to the (global) unbiased constraint:

$$\min \operatorname{var}(\hat{x}_i) \quad \text{s. t.} \quad \mathbf{e}_i^H \mathbf{h}_j = \delta_{i,j} \quad i, j = 1, \ldots, n \tag{106}$$

where $\mathbf{e}_i^H$ denotes the i^{th} row of the estimator matrix $\mathbf{E}$ and $\delta_{i,j}$ denotes the Kronecker delta. The constraints in (106) can also be compactly expressed as $\mathbf{EH} = \mathbf{I}$. 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 same constraint as in (106), such that the BLUE can also be interpreted as a Bayesian estimator. Similar arguments also hold for the 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 estimator 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

$$\hat{\mathbf{x}} = \mathbf{W}\underline{\mathbf{y}} = [\mathbf{E} \;\; \mathbf{F}]\underline{\mathbf{y}} = \mathbf{E}\mathbf{y} + \mathbf{F}\mathbf{y}^*. \tag{107}$$

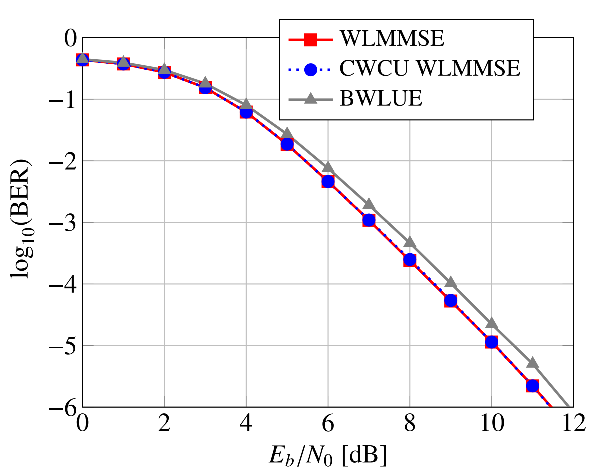

Fig. 5. Bit error ratio of different widely linear estimators for the described digital communication system setup in a multipath scenario. For the WLMMSE estimator, (104) was used for the LLR determination, while for the CWCU WLMMSE estimator the simpler expression in (105) was applied.

Table 2
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	$\operatorname{diag}\{\mathbf{EH}\} = \mathbf{1}$, $\mathbf{F} = \mathbf{0}$
BWLUE	$\mathbf{EH} = \mathbf{I}$, $\mathbf{FH}^* = \mathbf{0}$
WLMMSE	–
CWCU WLMMSE	$\operatorname{diag}\{\mathbf{EH}\} = \mathbf{1}$, $\operatorname{diag}\{\mathbf{FH}^*\} = \mathbf{0}$
CWCU WLMMSE for real parameter vectors	$\operatorname{diag}\{\mathbf{EH}\} + \operatorname{diag}\{\mathbf{FH}^*\} = \mathbf{1}$

Table 2 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 generally 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 generally 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 5.1).

7. 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 the CWCU WLMMSE estimator for a number of different preconditions, and started with jointly Gaussian parameters and measurements. Then, under linear model assumptions, we 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, we distinguished between improper complex and real parameters, which lead to different analytical expressions for the CWCU WLMMSE estimator. Finally, two examples where chosen to demonstrate the effects of the CWCU constraints and to highlight potential benefits of this type of estimator.

Acknowledgements

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

Appendix A. Derivation of the CWCU WLMMSE estimator for generalized jointly Gaussian x and y

In Appendix A we solve the optimization problem given in (55) using the Lagrange multiplier method. We start with the Lagrangian cost function which is

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

Using Wirtinger's calculus [43] for complex derivatives, we obtain

$$\frac{\partial J'}{\partial \mathbf{w}_i} = \underline{\mathbf{C}}_{\mathbf{yy}}^T \mathbf{w}_i^* + (\lambda^H \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}})^T. \tag{A.2}$$

By setting (A.2) equal to zero, $\mathbf{w}_i^H$ can be derived as

$$\mathbf{w}_i^H = -\lambda^H \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{A.3}$$

This result reinserted into the constraint in (55) leads to an expression for λ according to

$$\lambda^H = -[1 \;\; 0] \underline{\mathbf{C}}_{x_i x_i} (\underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y} x_i})^{-1} \underline{\mathbf{C}}_{x_i x_i}. \tag{A.4}$$

Eq. (A.4) reinserted into (A.3) leads to the final solution of the optimization problem in the form of

$$\mathbf{w}_{\mathrm{CWL},i}^H = [1 \;\; 0] \underline{\mathbf{C}}_{x_i x_i} (\underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y} x_i})^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{A.5}$$

Appendix B. Derivation of the conditional variance of the CWCU WLMMSE estimator

In Appendix B the conditional variance of the estimator regarded in Result 2 will be derived. For simplicity of the formulas we will denote $\mathbf{w}_{\mathrm{CWL},i}^H$ as $\mathbf{w}_i^H$ and $\hat{x}_{\mathrm{CWL},i}$ as $\hat{x}_i$, respectively. We then have

$$\mathrm{var}(\hat{x}_i | x_i) = E[|\hat{x}_i - E[\hat{x}_i | x_i]|^2 | x_i] = E[|\mathbf{w}_i^H \underline{\mathbf{y}} + b_i - \mathbf{w}_i^H E[\underline{\mathbf{y}} | x_i] - b_i|^2 | x_i] = E[|\mathbf{w}_i^H (\underline{\mathbf{y}} - E[\underline{\mathbf{y}} | x_i])|^2 | x_i] = \mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{yy}|x_i} \mathbf{w}_i. \tag{B.1}$$

Using

$$\underline{\mathbf{C}}_{\mathbf{yy}|x_i} = \underline{\mathbf{C}}_{\mathbf{yy}} - \underline{\mathbf{C}}_{\mathbf{y} x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{C}}_{x_i \mathbf{y}}, \tag{B.2}$$

which can be shown to hold for all three cases (a)–(c) in Result 2, together with (58) leads to

$$\mathrm{var}(\hat{x}_i | \mathrm{x}_i) = [1 \;\; 0] \underline{\mathbf{C}}_{x_i x_i} (\underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y} x_i})^{-1} \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - [1 \;\; 0] \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = [1 \;\; 0] \underline{\mathbf{C}}_{x_i x_i} (\underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y} x_i})^{-1} \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \sigma_{x_i}^2. \tag{B.3}$$

Appendix C. Derivation of the CWCU WLMMSE estimator for jointly Gaussian real x and $\mathbf{y}_{\mathbb{R}}$

In Appendix C we solve the optimization problem given in (91) using the Lagrange multiplier method. We start with the Lagrangian cost function which is

$$J' = \mathbf{w}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{w}_i - \sigma_{x_i}^2 + \lambda \left(\frac{1}{\sigma_{x_i}^2} [1 \;\; 0] \underline{\mathbf{C}}_{x_i \mathbf{y}} \mathbf{w}_i - 1 \right). \tag{C.1}$$

Using Wirtinger's calculus for complex derivatives, the derivation of (C.1) follows to

$$\frac{\partial J'}{\partial \mathbf{w}_i} = \underline{\mathbf{C}}_{\mathbf{yy}}^T \mathbf{w}_i^* + \left(\lambda \frac{1}{\sigma_{x_i}^2} [1 \;\; 0] \underline{\mathbf{C}}_{x_i \mathbf{y}} \right)^T. \tag{C.2}$$

By setting (C.2) equal to zero, $\mathbf{w}_i^H$ can be derived as

$$\mathbf{w}_i^H = -\lambda \frac{1}{\sigma_{x_i}^2} [1 \;\; 0] \underline{\mathbf{C}}_{x_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{C.3}$$

This result reinserted into the constraint in (91) leads to an expression for λ according to

$$\lambda = -\frac{(\sigma_{x_i}^2)^2}{[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}. \tag{C.4}$$

Eq. (C.4) reinserted into (C.3) leads to the final solution of the optimization problem in the form of

$$\mathbf{w}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{C.5}$$

Compounding $\mathbf{w}_{\mathrm{CWL},i}^H$ to an estimator matrix immediately leads to Result 3.

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

In Appendix D the conditional variance of the estimator regarded in Result 3 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 (87) and (92) we obtain

$$\mathrm{var}(\hat{x}_i|x_i) = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\mathbf{w}_i. \tag{D.1}$$

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

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

which can be shown to hold for all three cases (a)–(c) in Result 3. Multiplying (D.2) with $\mathbf{T}_n$ in (86) from the left and with $\mathbf{T}_n^H$ from the right yields

$$\underline{\mathbf{C}}_{\mathbf{yy}|x_i} = \underline{\mathbf{C}}_{\mathbf{yy}} - \mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\underline{\mathbf{y}}}. \tag{D.3}$$

Replacing $\mathbf{C}_{\underline{\mathbf{y}}x_i}$ with $\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}$ finally leads to

$$\underline{\mathbf{C}}_{\mathbf{yy}|x_i} = \underline{\mathbf{C}}_{\mathbf{yy}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2}[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}. \tag{D.4}$$

Inserting (D.4) into (D.1) results in

$$\mathrm{var}(\hat{x}_i|x_i) = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\mathbf{w}_i = \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{w}_i - \mathbf{w}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}\frac{1}{\sigma_{x_i}^2}[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\mathbf{w}_i = \frac{(\sigma_{x_i}^2)^2}{[1\ \ 0]\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}} - \sigma_{x_i}^2, \tag{D.5}$$

where in the last step, (92) has been incorporated.

References

[1] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, 1st edition, Prentice-Hall PTR, 2010 (April).

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

[3] Y.C. Eldar, Rethinking biased estimation improving maximum likelihood and the Cramér-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, Proceedings Asilomar Conference Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 2005, pp. 670–674.

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

[7] M. Triki, D.T.M. Slock, Investigation of some bias and MSE issues in block-component-wise conditionally unbiased LMMSE, Proceedings Asilomar Conference Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 2006, pp. 1420–1424.

[8] O. Lang, M. Huemer, On the Log-Likelihood Ratio Evaluation of CWCU Linear and Widely Linear MMSE Data Estimators, accepted for publication in the Proceedings Asilomar Conference Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 2016.

[9] M. Huemer, O. Lang, On component-wise conditionally unbiased linear Bayesian estimation, Proceedings Asilomar Conference Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 2014, pp. 879–885.

[10] 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, Proceedings of the 15th International Conference, Las Palmas de Gran Canaria, Spain, Feb. 2015, revised selected papers, vol. 9520, December 2015. pp. 537–545.

[11] T. Adali, P.J. Schreier, L.L. Scharf, Complex-valued signal processing: the proper way to deal with impropriety, IEEE Trans. Signal Process. 59 (11) (2011) 5101–5125.

[12] B. Picinbono, P. Chevalier, Widely linear estimation with complex data, IEEE Trans. Signal Process. 43 (8) (1995) 2030–2033.

[13] P.J. Schreier, L.L. Scharf, Statistical Signal Processing of Complex-valued Data: The Theory of Improper and Noncircular Signals, Cambridge University Press, 2010.

[14] M. Witzke, Linear and widely linear filtering applied to iterative detection of generalized MIMO signals, Ann. Telecommun. 60 (2005) 147–168.

[15] J. Via, D. Ramirez, I. Santamaria, L. Vielva, Widely and semiwidely linear processing of quaternion vectors, Proceedings IEEE International Conference Acoust., Speech, Signal Process, Dallas, TX, USA, Mar. 2010, pp. 3946–3949.

[16] P. Chevalier, J.P. Delmas, A. Oukaci, Optimal widely linear MVDR beamforming for noncircular signals, Proceedings International Conference Acoust., Speech, Signal Process, Taipei, Taiwan, Apr. 2009, pp. 3573–3576.

[17] A. Mirbagheri, N. Plataniotis, S. Pasupathy, An enhanced widely linear CDMA receiver with OQPSK modulation, IEEE Trans. Commun. 54 (2006) 261–272.

[18] C. Stanciu, J. Benesty, C. Paleologu, T. Gnsler, S. Ciochin, A widely linear model for stereophonic acoustic echo cancellation, Signal Process. 93 (February (2)) (2013) 511–516.

[19] D. Xu, C. Gong, S. Cao, X. Xu, Z. Ye, Robust widely linear beamforming based on spatial spectrum of noncircularity coefficient, Signal Process. 104 (2014) 167–173.

[20] P. Chevalier, A. Blin, Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences, IEEE Trans. Signal Process 55 (11) (2007) 5323–5336.

[21] F. Sterle, D. Mattera, L.Paura, Two nonequivalent structures for widely-linear decision-feedback MMSE equalization over MIMO channels, Proceedings of the 14th European Signal Processing Conference, Florence, Italia, Sep. 2006. pp. 1–5.

[22] K. Zeng, Y. Tang, S. Li, Widely linear decision-feedback detection for distributed transmit antenna MIMO systems, Proceedings of the IET International Conference on Wireless, Mobile and Multimedia Networks, Hangzhou, China, Nov. 2006, pp. 1–4.

[23] D. Mattera, L. Paura, F. Sterle, Widely linear decision-feedback equalizer for time-dispersive linear MIMO channels, IEEE Trans. Signal Process 53 (7) (2005) 2525–2536.

[24] C. Jahanchahi, S. Kanna, D. Mandic, Complex dual channel estimation: cost effective widely linear adaptive filtering, Signal Process. 104 (2014) 33–42.

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

[26] E. Ollila, On the circularity of a complex random variable, IEEE Signal Process Lett. 15 (2008) 841–844.

[27] M. Novey, T. Adali, A. Roy, Circularity and Gaussianity detection using the complex generalized Gaussian distribution, EEE Signal Proceedings Lett. 16 (1) (2009) 993–996.

[28] J. Eriksson, E. Ollila, V. Koivunen, Essential statistics and tools for complex random variables, IEEE Trans. Signal Process 58 (10) (2010) 5400–5408.

[29] A.T. Walden, P. Rubin-Delanchy, On testing for impropriety of complex-valued Gaussian vectors, IEEE Trans. Signal Process 57 (3) (2009) 825–834.

[30] P.J. Schreier, L.L. Scharf, Second-order analysis of improper complex random vectors and processes, IEEE Trans. Signal Process 51 (3) (2003) 714–725.

[31] B. Picinbono, On circularity, IEEE Trans. Signal Process 42 (12) (1994) 3473–3482.

[32] M. Novey, T. Adali, A. Roy, A complex generalized Gaussian distribution characterization, generation, and estimation, IEEE Trans. Signal Process 58 (3) (2010) 1427–1433.

[33] B. Picinbono, Second-order complex random vectors and normal distributions, IEEE Trans. Signal Process 44 (10) (1996) 2637–2640.

[34] A. Van Den Bos, The multivariate complex normal distribution-a generalization, IEEE Trans. Inf. Theory 41 (2) (1995) 537–539.

[35] M. Huemer, C. Hofbauer, J.B. Huber, Non-systematic complex number RS coded OFDM by unique word prefix, IEEE Trans. Signal Process 60 (1) (2012) 285–299.

[36] M. Huemer, C. Hofbauer, A. Onic, J.B. Huber, Design and analysis of UW-OFDM signals, AEU - Int. J. Electron. Commun. 68 (10) (2014) 958–968.

[37] IEEE Std 802.11a-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-Speed Physical Layer in the 5 GHz Band, 1999.

[38] S. Allpress, C. Luschi, S. Felix, Exact and approximated expressions of the log-likelihood ratio for 16-QAM signals, Proceedings Asilomar Conference Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 2004, pp. 794–798.

[39] J. Fakatselis, Criteria for 2.4 GHz PHY Comparison of Modulation Methods, Document IEEE 1997; P802.11-97/157r1.

[40] Ki Seol Kim, Kwangmin Hyun, Chang Wahn Yu, Youn Ok Park, Dongweon Yoon, Sang Kyu Park, General log-likelihood ratio expression and its implementation algorithm for gray-coded QAM signals, ETRI J. 28 (3) (2006) 291–300 (June).

[41] Kyongkuk Cho, Dongweon Yoon, On the general BER expression of one- and two-dimensional amplitude modulations, IEEE Trans. Commun. 50 (7) (2002) 1074–1080 (July).

[42] P. Robertson, P. Hoeher, E. Villebrun, Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding, Eur. Trans. Telecommun. (ETT) 8 (119–125) (1997) (March).

[43] W. Wirtinger, Zur formalen Theorie der Funktionen von mehr komplexen Veränderliche, Math. Ann. 97 (1927) 357–375.