JOHANNES KEPLER
UNIVERSITY LINZ

Submitted by
Oliver Lang

Submitted at
Institute of Signal Processing

Supervisor and First Examiner
Univ.-Prof. Dr. Mario Huemer

Second Examiner
Prof. Dr. Stephan Weiss

March 2018

Knowledge-Aided Methods in Estimation Theory and Adaptive Filtering

Doctoral Thesis
to obtain the academic degree of
Doktor der technischen Wissenschaften
in the Doctoral Program
Technische Wissenschaften

JOHANNES KEPLER UNIVERSITY LINZ
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

Abstract

Estimation theory is a key enabler in many of today's electronic products, devices, and industrial equipment. Among others, it provides the basis for efficient data estimation in communication systems, accurate characterization of systems based on measurements, estimation of parameters, signals and spectra, signal tracking, or noise cancellation, to name just a few. The estimation task can be described in a classical or in a Bayesian framework. In classical estimation the parameter vector to be estimated is considered to be deterministic. Conversely, Bayesian estimators consider the parameter vector to be random. This allows to include prior knowledge in form of statistics of the parameter vector into the estimation problem.

Due to the ever-increasing complexity and the more demanding applications of modern electronic systems, optimal or near-to-optimal performance of the estimation methods is often required. To achieve such an optimal performance, every available information about the underlying system model should be incorporated by the estimators. Ultimately, however, additional model knowledge is present in many applications. This model knowledge is often ignored when developing the estimators. Possible examples of additional model knowledge are:

- the knowledge that the parameter vector of length n lies in a linear subspace of $\mathbb{C}^n$,
- the knowledge that the parameter vector fulfills additional linear constraints,
- the knowledge that the parameter vector is real-valued while the measurements and the measurement noise are complex-valued,
- and the knowledge that the measurement matrix is subject to an unknown random error with known second order statistics.

For the first three cases, several *knowledge-aided* classical estimators are proposed in this thesis that incorporate the available model knowledge in an optimal way. Simulation examples are presented demonstrating the performance gain of the derived estimators compared to state-of-the-art estimators. Moreover, the derived estimators also are compared to intuitive estimators that incorporate the additional model knowledge in an intuitive manner. It turns out that the derived optimal estimators significantly outperform these intuitive estimators as well as state-of-the-art estimators in many scenarios. For the fourth case of additional model knowledge, a novel iterative algorithm is proposed. It is shown that this algorithm outperforms competing algorithms significantly in many scenarios.

Another difference between the classical and Bayesian approaches is the considered unbiased constraint. We discuss the fact that the unbiased constraint utilized by state-of-the-art Bayesian estimators is weaker than that utilized by unbiased classical estimators. We furthermore show that this weaker unbiased constraint is the key enabler for Bayesian estimators to incorporate statistics about the unknown parameter vector into the estimation process. Based on that, we investigate the so called component-wise conditionally unbiasedness constraints. It will be shown, that these unbiased constraints preserve the intuitive view of unbiasedness also in Bayesian scenarios. Next, this thesis focusses on the class of so called component-wise conditionally unbiased Bayesian estimators. We will extend previous work on this type of estimator and extend the concept to widely linear estimators. The effects of these unbiased constraints, the relation to other Bayesian estimators and the ability to incorporate statistics about the unknown parameter vector are discussed.

Based on the performance gain achievable by classical estimators incorporating additional model knowledge, we derive adaptive filters that also incorporate such model knowledge in an optimal way. These knowledge-aided adaptive filters are compared with intuitive as well as state-of-the-art adaptive filters, where again a significant performance boost is achieved in many scenarios. Furthermore, adaptive filters for the task of system identification are developed that allow incorporating prior knowledge about the impulse response of the system. Existing and newly proposed adaptive filters utilizing prior knowledge are discussed and compared.

Kurzfassung

Die Schätztheorie ist ein Schlüsselfaktor für viele der heutigen elektronischen Produkte, Geräte und Industrieanlagen. Unter anderem stellt diese Algorithmen zur effizienten Datenschätzung in Kommunikationssystemen, zur genauen Charakterisierung von Systemen basierend auf Messungen, Schätzung von Parametern, Signalen und Spektren, Signalverfolgung oder Rauschunterdrückung, zur Verfügung, um nur einige zu nennen. Die Schätzaufgabe kann in einem klassischen oder in einem Bayes'schen Rahmen formuliert werden. In der klassischen Schätzung wird der zu schätzende Parametervektor als deterministisch angesehen. Im Gegensatz dazu betrachten Bayes'sche Schätzer den Parametervektor als zufällig. Dies ermöglicht es, Vorkenntnisse in Form von Statistiken des Parametervektors in das Schätzproblem einzubeziehen.

Aufgrund der ständig zunehmenden Komplexität und der anspruchsvolleren Anwendungen moderner elektronischer Systeme ist oft eine optimale oder nahezu optimale Performance der Schätzverfahren erforderlich. Um eine solche optimale Performance zu erzielen sollten alle verfügbaren Informationen über das zugrundeliegende Systemmodell von den Schätzern einbezogen werden. In vielen Anwendungen ist tatsächlich zusätzliches Modellwissen vorhanden. Dieses Modellwissen wird bei der Entwicklung der Schätzer jedoch oft ignoriert. Mögliche Beispiele für zusätzliches Modellwissen sind die Kenntnis,

- dass der Parametervektor der Länge n in einem linearen Unterraum von $\mathbb{C}^n$ liegt,
- dass der Parametervektor zusätzliche lineare Bedingungen erfüllt,
- dass der Parametervektor reellwertig ist während die Messungen und das Messrauschen komplexwertig sind,
- dass die Verbindung zwischen den Messungen und den Parametern durch Messrauschen sowie durch eine zufällige Verzerrung mit bekannten Statistiken beeinflusst wird.

Für die ersten drei der oben genannten Fälle werden in dieser Arbeit mehrere wissensunterstützte klassische Schätzer entwickelt, die dieses zusätzliche Modellwissen optimal verarbeiten. Diese optimalen wissensunterstützten Schätzer werden mit Schätzern verglichen, die das zusätzliche Modellwissen intuitiv verarbeiten. Es stellt sich heraus, dass die hergeleiteten optimalen Schätzer die intuitiven Schätzer und Standard-Schätzer in vielen Anwendungen deutlich in ihrer Performance übertreffen. Für den vierten Fall von zusätzlichem Modellwissen wird ein neuer iterativer Algorithmus hergeleitet. Es wird gezeigt, dass dieser Algorithmus konkurrierende Algorithmen in vielen Szenarien deutlich in der Schätzgenauigkeit übertrifft.

Ein weiterer Unterschied zwischen dem klassischen und dem Bayes'schen Ansatz ist die zugrundeliegende Definition eines erwartungstreuen Schätzers. Wir diskutieren die Tatsache, dass die Bedingung der Erwartungstreue, die von Bayes'schen Schätzern verwen-

det wird, schwächer ist als die, die von erwartungstreuen klassischen Schätzern verwendet wird. Wir zeigen außerdem, dass diese schwächere Bedingung der Erwartungstreue der Schlüssel dafür ist, dass Bayes'sche Schätzer Statistiken über den unbekannten Parametervektor in den Schätzprozess einbeziehen können. Darauf aufbauend untersuchen wir Bedingungen für die sogenannte komponentenweise bedingte Erwartungstreue (engl.: component-wise conditionally unbiased (CWCU) constraints). Es wird gezeigt, dass die zugrundeliegenden CWCU Bedingungen die intuitive Sicht der Erwartungstreue auch in Bayes'schen Szenarien bewahren. Als nächstes konzentrieren wir uns die Gruppe der sogenannten CWCU Bayes'schen Schätzer. Wir werden bisherige Arbeiten zu dieser Art von Schätzern erweitern und das Konzept auf sogenannte *widely linear* Schätzer ausweiten. Die Auswirkungen dieser CWCU Bedingungen, die Beziehung zu anderen Bayes'schen Schätzern und die Fähigkeit, Statistiken über den unbekannten Parametervektor einzubauen, werden diskutiert.

Basierend auf der erhöhten Schätzgenauigkeit die durch klassische Schätzer erreicht werden kann welche zusätzliches Modellwissen nutzen werden weiters adaptive Filter hergeleitet, die ebenfalls zusätzliches Modellwissen auf optimale Weise einbeziehen. Diese wissensunterstützten adaptiven Filter werden sowohl mit intuitiv entwickelten Filtern als auch mit adaptiven Standard-Filter verglichen, wobei in vielen Szenarien wiederum eine deutliche Erhöhung der Schätzgenauigkeit erreicht wird. Darüber hinaus werden adaptive Filter für System-Identifikations-Anwendungen untersucht die es erlauben ähnliche statistische Vorkenntnisse über die zu schätzende Systemimpulsantwort einzubringen, wie dies bei linearen Bayes'schen Schätzern der Fall ist. Bekannte und neu entwickelte adaptive Filter werden diskutiert und verglichen.

Statutory Declaration

I hereby declare that the thesis submitted is my own unaided work, that I have not used other than the sources indicated, and that all direct and indirect sources are acknowledged as references.
This printed thesis is identical with the electronic version submitted.

Date	Signature

Acknowledgements

There are many people for whom I would like to express my thanks. Most of all, I would like to thank my advisor, Univ.-Prof. Dr. Mario Huemer, for entrusting me with the position as university assistant at the Institute of Signal Processing. Working with you was very educational and inspiring and I will always appreciate your mentorship and support during challenging times at the Institute.

I would also like to thank my colleagues at the Institute of Signal Processing for the great time we spent together. You have created an environment in which I felt comfortable and encouraged. I will keep many great memories.

And finally, I want to thank my family for all their support ever since I can remember.

Contents

1 Introduction

In signal processing, estimation is the task of approximating meaningful values that are useful for further processing, for describing a system, or some other purpose [1]. These values are usually called parameters or states, depending on the context. The estimates of the unknown parameters are typically derived based on known measurements. Most estimation techniques require knowledge about the connection between the measurements and the parameters, which is usually described by the model. The most common of all models is the linear model, where the measurements are linearly connected to the parameters. Estimation tasks with an underlying linear model appear in a very broad band of technical fields, e.g., radar, communications, control, biomedical engineering, image and speech analysis. Many facts about estimation with an underlying linear model are already derived and available in standard literature such as [1]. However, there still exist some novel and exciting aspects that deserve investigation.

The estimation task can be done in different contexts [1]. The classical context (also known as the frequentist context) treats the parameters as unknown but deterministic. Here, the performance criterion is in most cases the mean square error (MSE) between the estimated and true parameters, averaged over the probability density function (PDF) of the measurements. The Bayesian context on the other hand treats the unknown parameters as random variables whose particular realizations have to be estimated. This approach allows assigning statistics or even a full PDF to the parameters since they represent random variables. These quantities are termed *prior knowledge*. By incorporating this prior knowledge into the estimation process, the performance in terms of the MSE may be improved significantly. It turns out, however, that in general the MSE performance depends on the actual realization of the parameters [1]. To obtain a performance measure that is independent of the particular realization of the parameters, the Bayesian mean square error (BMSE) is usually utilized. It corresponds to the MSE when averaged over the PDF of the parameters.

Let us consider the linear model, which describes the linear connection between the measurements and the parameters. These measurements as well as the parameters may be considered real- or complex-valued. In this work, almost all investigations are carried out for complex-valued quantities. Exceptions are mentioned explicitly. Complex-valued models and signals appear in many technical areas. Prominent examples for complex-valued signals are baseband signals in radar or communication applications. Some widely used transformations that utilize complex-valued numbers are the Hilbert transform and the fast Fourier transform (FFT) [2]. To represent complex-valued vectors and matrices, we use the so-called augmented notation [3]. This representation has the advantage that the results often appear in compact form. Moreover, structural similarities with the expressions that occur for real-valued models and signals become apparent.

The outline of this thesis is as follows. Chapter 2 recapitulates the fundamental mathematical concepts required in this work. There, the augmented notation is described as well as an introduction to the Wirtinger calculus [4], and the Lagrange multiplier method for the case of complex-valued variables [5,6] is provided. The novelty of this work is presented in Chapters 3, 4, and 5. The focus of these three parts are knowledge-aided concepts in classical estimation (Chapter 3), a rarely investigated Bayesian estimator (Chapter 4), and knowledge-aided concepts in adaptive filtering (Chapter 5), respectively. In the following, the main investigations of these three chapters are summarized. Details as well as a discussion of the current state-of-the-art are provided at the beginning of each chapter.

Knowledge-Aided Concepts in Classical Estimation

Chapter 3 starts by briefly reviewing standard classical estimators like the least squares (LS) estimator, the best linear unbiased estimator (BLUE) [1] and the best widely linear unbiased estimator (BWLUE) [2] for the linear model case. For these estimators we then rigorously regard the commutation analysis over linear transformations. Most of these properties can be found in standard literature, however, we find that the issue is sometimes treated superficially in engineering literature.

The linear model in many cases implies that the measurement matrix and sometimes the covariance matrix of the noise are known. Both quantities are incorporated, e.g., by the well-known classical BLUE. The main part of Chapter 3 deals with the investigation and derivation of classical estimators that use additional model knowledge that might be available in practice. The resulting optimal knowledge-aided estimators are compared with competing standard and intuitive estimators. Four cases of additional model knowledge are considered.

The first case of additional model knowledge is the knowledge that the parameter vector of length n lies in a linear subspace of $\mathbb{C}^n$. It is proven that standard classical estimators such as the BLUE and the BWLUE can incorporate this additional knowledge in a straightforward manner. On the other hand, for the LS estimator it is shown in this thesis that a constrained LS estimator [1] is able to incorporate the additional model knowledge. The linear constraints required for applying the constrained LS estimator are derived.

Secondly, the knowledge that the parameter vector fulfills additional linear constraints is considered. In that case, the constrained LS estimator is available as a standard estimator. For the BLUE and the BWLUE, no corresponding extension exists in the literature to the best of our knowledge. This gap is closed by proposing the constrained

BLUE and the constrained BWLUE. These novel estimators incorporate the fact that the parameter vector fulfills additional linear constraints allowing to increase the estimation accuracy compared to the standard BLUE and BWLUE.

Thirdly, we regard problems for which it is known that the parameters are real-valued while the measurements are complex-valued. If this is the case, applying the ordinary BLUE in general results in complex-valued estimates, producing a systematic error. In order to prevent this systematic error, several novel classical estimators are proposed that incorporate this additional model knowledge in an optimal way. We show that by incorporating the knowledge that the true parameters are real-valued, the estimator's performance can be increased significantly.

The fourth investigated case of additional model knowledge considers estimation tasks, where the measurement matrix is not completely known but is subject to errors with known error variances [7–12]. Typical practical applications are problems for which the measurement matrix is e.g. a convolution matrix that is constructed based on an imperfectly measured or estimated impulse response. Incorporating the error variances into the estimation process allows to significantly increasing the estimation accuracy. In this thesis, an iterative estimation algorithm is proposed that outperforms existing algorithms.

For the first three mentioned cases of additional model knowledge, we derive optimal estimators and compare them with standard estimators as well as trivial estimators that incorporate the additional model knowledge in an intuitive way. For the forth case optimality cannot be claimed, however, it is shown that the proposed algorithm outperforms competing algorithms by far. Note that all mentioned cases of additional model knowledge, if appropriate, usually directly follow from the physical circumstances of the underlying problem. This makes this knowledge easy available once the model is known. A summary of practical applications complete these investigations.

Component-Wise Conditionally Unbiased LMMSE and WLMMSE Estimation

Chapter 4 deals with a particular class of Bayesian estimators, the so-called component-wise conditionally unbiased (CWCU) estimators. In contrast to classical estimators, Bayesian estimators consider the parameter vector to be random with known statistics. These statistics are termed prior knowledge. This prior knowledge in many cases allows to significantly improve the estimation accuracy compared to classical estimators. Another difference between classical unbiased and typical Bayesian estimators is the arising unbiased constraint. The unbiased constraint utilized by state-of-the-art Bayesian estimators is weaker than that utilized by unbiased classical estimators. In fact, the linear minimum mean square error (LMMSE) estimator and the widely linear minimum mean square error (WLMMSE) estimator are conditionally biased. In light of this, we investigate the CWCU constraints. It is shown that estimators fulfilling these constraints in many cases also allow to incorporate prior knowledge into the estimation process. Along with this, the intuitive view of classical unbiasedness is preserved in Bayesian scenarios. Chapter 4 extends the results on component-wise conditionally unbiased linear minimum mean square error (CWCU LMMSE) estimators as derived in [13–15], and the component-wise conditionally unbiased widely linear minimum mean square error (CWCU WLMMSE) estimator is introduced and derived for complex-valued parameters as well as for real-valued parameters and complex-valued measurements. It will be shown that the derived CWCU estimators are closely related to their LMMSE and WLMMSE counterparts, but avoid some of their typical effects. These effects will be discussed and examples where CWCU estimators can be beneficially employed are presented.

Knowledge-Aided Concepts in Adaptive Filtering

In Chapter 5, the well-known least mean square (LMS) and recursive least squares (RLS) adaptive filter algorithms are recapitulated and extended. These extensions can be separated into two major parts.

In the first part, we incorporate the additional model knowledge, that the true filter coefficients should be real-valued while the input and desired signal are complex-valued. A practical example where this situation can arise is given in the simulation section of this chapter, where the problem of transmit leakage in modern wireless transceivers is considered. One way to extract and cancel the leakage signal is to use a so called auxiliary receiver in parallel to the main receiver [16]. In such an application, however, typically a fractional delay between the two receivers appears. Adaptive filters can be used to estimate and compensate for this fractional delay. In this application the input and desired signal are complex-valued while the optimum filter coefficients are real-valued. In this thesis, novel extensions of the LMS and RLS algorithms that use this additional model knowledge in an optimal way and that produce real-valued filter coefficients are developed. These optimal filters are compared with state-of-the-art filters as well as with trivial filters that incorporate the additional model knowledge in an intuitive way. In the case of the LMS algorithm it turns out that the intuitive filter corresponds to the optimal approach. In case of the RLS algorithm, however, the derived optimal algorithm outperforms the intuitive filter significantly as will be demonstrated in simulation examples.

In the second part, the system identification application of adaptive filters is considered. In many applications prior statistical knowledge about the impulse response to be estimated is available. An adaptive filter that is able to incorporate prior knowledge and that is related to the RLS algorithm is the sequential LMMSE estimator in a filtering setup. A similar extension based on the LMS algorithm has been derived in the context of this doctoral thesis work. This algorithm allows to incorporate the first and second order statistical moments about the impulse response to be estimated. It is shown that these adaptive filters incorporating prior knowledge are able to reduce the convergence time in the mean compared to their standard LMS and RLS counterparts.

Major findings presented in this work are marked as *Results*. These results are either novel and unpublished to the best of our knowledge, or published by the author of this work himself. A small frame around them visually emphasizes the results. The simulation examples are marked with a black bar on the left side.

2

Prerequisites

This chapter summarizes the prerequisites required in the remainder of this work. We begin with the notation and fundamental definitions.

2.1 Notation

Lower-case bold face variables ($\mathbf{a}$, $\mathbf{b}$,...) indicate vectors, and upper-case bold face variables ($\mathbf{A}$, $\mathbf{B}$,...) indicate matrices. We further use $\mathbb{R}$ and $\mathbb{C}$ to denote the set of real and complex numbers, respectively, $(\cdot)^*$ to denote the complex conjugate, $(\cdot)^T$ to denote transposition, $(\cdot)^H$ 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 $\mathbf{0}$. The subscript $_\mathrm{R}$ of a vector or matrix denotes its real part and the subscript $_\mathrm{I}$ 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.

2.2 Augmented Form and Widely Linear Processing

This section recapitulates the preliminaries required to derive the linear and particularly the widely linear estimators in this work. It is essentially a shortened version of the corresponding parts in [2,3], where an excellent introduction to improper data and widely linear processing can be found. It will turn out, that widely linear processing allows to incorporate improper statistics into the estimation process, while standard complex-valued processing only allows to incorporate proper statistics.

2.2.1 Linear and Widely Linear Transformations

We write a *complex vector* $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ as $\mathbf{x} = \mathbf{x}_\mathrm{R} + j\mathbf{x}_\mathrm{I}$, where $\mathbf{x}_\mathrm{R} = \mathrm{Re}\{\mathbf{x}\} \in \mathbb{R}^{N_\mathbf{x}}$ and $\mathbf{x}_\mathrm{I} = \mathrm{Im}\{\mathbf{x}\} \in \mathbb{R}^{N_\mathbf{x}}$. Based on that, we use two closely related representations. The first representation is the *real composite* $2N_{\mathbf{x}}$-dimensional vector

$$\mathbf{x}_{\mathbb{R}} = \begin{bmatrix} \mathbf{x}_{\mathrm{R}} \\ \mathbf{x}_{\mathrm{I}} \end{bmatrix} \in \mathbb{R}^{2N_{\mathbf{x}}}, \tag{2.1}$$

obtained by stacking $\mathbf{x}_{\mathrm{R}}$ on top of $\mathbf{x}_{\mathrm{I}}$. The second representation is the *complex augmented vector*

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix}, \tag{2.2}$$

obtained by stacking $\mathbf{x}$ on top of its complex conjugate $\mathbf{x}^*$. Augmented vectors are always underlined. In much of our discussion, our focus will be on complex-valued quantities, where we will be using $\mathbf{x}$ and its augmentation $\underline{\mathbf{x}}$.

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

$$\mathbf{T}_{N_{\mathbf{x}}} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \in \mathbb{C}^{2N_{\mathbf{x}}\times 2N_{\mathbf{x}}} \tag{2.3}$$

is unitary up to a factor of 2, i.e., $\mathbf{T}_{N_{\mathbf{x}}}\mathbf{T}_{N_{\mathbf{x}}}^H = \mathbf{T}_{N_{\mathbf{x}}}^H\mathbf{T}_{N_{\mathbf{x}}} = 2\mathbf{I}$. The complex augmented vector $\underline{\mathbf{x}}$ is obviously an equivalent redundant, but convenient representation of $\mathbf{x}_{\mathbb{R}}$.

In the following, we consider widely linear transformations of the form

$$\mathbf{y} = \mathbf{H}_1\mathbf{x} + \mathbf{H}_2\mathbf{x}^*. \tag{2.4}$$

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

$$\underline{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^* & \mathbf{H}_1^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix} = \underline{\mathbf{H}}\,\underline{\mathbf{x}}. \tag{2.5}$$

The matrix $\underline{\mathbf{H}}$ is called an *augmented matrix*. It satisfies a particular block pattern, where the south-east block is the conjugate of the north-west block, and where the south-west block is the conjugate of the north-east block.

We now apply the concept of complex augmented vectors and matrices on the linear model given by

$$\mathbf{y} = \mathbf{H}\,\mathbf{x} + \mathbf{n}, \tag{2.6}$$

where $\mathbf{H} \in \mathbb{C}^{N_{\mathbf{y}}\times N_{\mathbf{x}}}$ is a known measurement matrix, $\mathbf{x}$ is the unknown parameter vector, $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ is the measurement vector, and $\mathbf{n} \in \mathbb{C}^{N_{\mathbf{y}}}$ is a zero mean noise vector statistically independent of $\mathbf{x}$. The augmented version of (2.6) is

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{n}}, \tag{2.7}$$

where

$$\underline{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^* \end{bmatrix}, \quad \underline{\mathbf{n}} = \begin{bmatrix} \mathbf{n} \\ \mathbf{n}^* \end{bmatrix}. \tag{2.8}$$

2.2.2 Statistics of Complex-Valued Random Vectors

In order to characterize the second-order statistical properties of $\mathbf{x} = \mathbf{x}_\mathrm{R} + j\mathbf{x}_\mathrm{I}$, we start by considering the real composite random vector $\mathbf{x}_\mathbb{R} = \begin{bmatrix}\mathbf{x}_\mathrm{R}^T & \mathbf{x}_\mathrm{I}^T\end{bmatrix}^T$. Its covariance matrix is

$$\mathbf{C}_{\mathbf{x}_\mathbb{R}\mathbf{x}_\mathbb{R}} = E_{\mathbf{x}_\mathbb{R}}[(\mathbf{x}_\mathbb{R} - E_{\mathbf{x}_\mathbb{R}}[\mathbf{x}_\mathbb{R}])(\mathbf{x}_\mathbb{R} - E_{\mathbf{x}_\mathbb{R}}[\mathbf{x}_\mathbb{R}])^T] = \begin{bmatrix}\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} & \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}} \\ \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}^T & \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}}\end{bmatrix} \tag{2.9}$$

with

$$\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} = E_{\mathbf{x}_\mathrm{R}}[(\mathbf{x}_\mathrm{R} - E_{\mathbf{x}_\mathrm{R}}[\mathbf{x}_\mathrm{R}])(\mathbf{x}_\mathrm{R} - E_{\mathbf{x}_\mathrm{R}}[\mathbf{x}_\mathrm{R}])^T] \tag{2.10}$$

$$\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}} = E_{\mathbf{x}_\mathrm{R},\mathbf{x}_\mathrm{I}}[(\mathbf{x}_\mathrm{R} - E_{\mathbf{x}_\mathrm{R}}[\mathbf{x}_\mathrm{R}])(\mathbf{x}_\mathrm{I} - E_{\mathbf{x}_\mathrm{I}}[\mathbf{x}_\mathrm{I}])^T] \tag{2.11}$$

$$\mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}} = E_{\mathbf{x}_\mathrm{I}}[(\mathbf{x}_\mathrm{I} - E_{\mathbf{x}_\mathrm{I}}[\mathbf{x}_\mathrm{I}])(\mathbf{x}_\mathrm{I} - E_{\mathbf{x}_\mathrm{I}}[\mathbf{x}_\mathrm{I}])^T]. \tag{2.12}$$

The augmented covariance matrix of $\mathbf{x}$ is defined as

$$\underline{\mathbf{C}}_{\mathbf{xx}} = E_\mathbf{x}[(\underline{\mathbf{x}} - E_\mathbf{x}[\underline{\mathbf{x}}])(\underline{\mathbf{x}} - E_\mathbf{x}[\underline{\mathbf{x}}])^H]. \tag{2.13}$$

With the real-to-complex transformation matrix $\mathbf{T}_{N_\mathbf{x}}$ we have

$$\underline{\mathbf{C}}_{\mathbf{xx}} = \mathbf{T}_{N_\mathbf{x}}\mathbf{C}_{\mathbf{x}_\mathbb{R}\mathbf{x}_\mathbb{R}}\mathbf{T}_{N_\mathbf{x}}^H \tag{2.14}$$

$$= \begin{bmatrix}\mathbf{C}_{\mathbf{xx}} & \widetilde{\mathbf{C}}_{\mathbf{xx}} \\ \widetilde{\mathbf{C}}_{\mathbf{xx}}^* & \mathbf{C}_{\mathbf{xx}}^*\end{bmatrix} = \underline{\mathbf{C}}_{\mathbf{xx}}^H \in \mathbb{C}^{2N_\mathbf{x}\times 2N_\mathbf{x}}, \tag{2.15}$$

where

$$\mathbf{C}_{\mathbf{xx}} = E_\mathbf{x}[(\mathbf{x} - E_\mathbf{x}[\mathbf{x}])(\mathbf{x} - E_\mathbf{x}[\mathbf{x}])^H] \tag{2.16}$$

is the (Hermitian and positive semi-definite) covariance matrix, and where

$$\widetilde{\mathbf{C}}_{\mathbf{xx}} = E_\mathbf{x}[(\mathbf{x} - E_\mathbf{x}[\mathbf{x}])(\mathbf{x} - E_\mathbf{x}[\mathbf{x}])^T] \tag{2.17}$$

is the complementary covariance matrix. An equivalent expression to (2.14) is

$$\mathbf{C}_{\mathbf{x}_\mathbb{R}\mathbf{x}_\mathbb{R}} = = \frac{1}{4}\mathbf{T}_{N_\mathbf{x}}^H\underline{\mathbf{C}}_{\mathbf{xx}}\mathbf{T}_{N_\mathbf{x}}. \tag{2.18}$$

For $\mathbf{C}_{\mathbf{xx}}$ and $\widetilde{\mathbf{C}}_{\mathbf{xx}}$ we have

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} + \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}} + j(\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}^T - \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}) = \mathbf{C}_{\mathbf{xx}}^H, \tag{2.19}$$

and

$$\widetilde{\mathbf{C}}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} - \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}} + j(\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}^T + \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}) = \widetilde{\mathbf{C}}_{\mathbf{xx}}^T, \tag{2.20}$$

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* [17–21]. The conditions for propriety on the covariance and cross-covariance of real and imaginary parts $\mathbf{x}_\mathrm{R}$ and $\mathbf{x}_\mathrm{I}$ are $\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} = \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}}$ and $\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}} = -\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}}^T$. When $x =$

$x_{\mathrm{R}} + jx_{\mathrm{I}}$ is scalar, then $C_{x_{\mathrm{R}}x_{\mathrm{I}}} = 0$ is necessary for propriety. If $\mathbf{x}$ is proper, its Hermitian covariance matrix is

$$\mathbf{C}_{\mathbf{xx}} = 2\mathbf{C}_{\mathbf{x}_{\mathrm{R}}\mathbf{x}_{\mathrm{R}}} - 2j\mathbf{C}_{\mathbf{x}_{\mathrm{R}}\mathbf{x}_{\mathrm{I}}} = 2\mathbf{C}_{\mathbf{x}_{\mathrm{I}}\mathbf{x}_{\mathrm{I}}} + 2j\mathbf{C}^{T}_{\mathbf{x}_{\mathrm{R}}\mathbf{x}_{\mathrm{I}}}, \tag{2.21}$$

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

2.2.3 Linear and Widely Linear Estimators

Let $\mathbf{x} \in \mathbb{C}^{N_{\mathbf{x}}}$ be the parameter vector to be estimated and $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ be the measurement vector. Then, a widely linear (or actually widely affine) estimator takes on the form

$$\hat{\mathbf{x}} = \mathbf{F}\mathbf{y} + \mathbf{G}\mathbf{y}^* + \mathbf{b}, \tag{2.22}$$

where $\mathbf{F}$, $\mathbf{G} \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\mathbf{y}}}$ and $\mathbf{b} \in \mathbb{C}^{N_{\mathbf{x}}}$. Another way to express the estimator is its augmented version

$$\underline{\hat{\mathbf{x}}} = \begin{bmatrix} \mathbf{F} & \mathbf{G} \\ \mathbf{G}^* & \mathbf{F}^* \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} + \underline{\mathbf{b}} = \underline{\mathbf{E}}\underline{\mathbf{y}} + \underline{\mathbf{b}}. \tag{2.23}$$

For linear (or actually affine) estimators we have $\mathbf{G} = \mathbf{0}$ such that $\hat{\mathbf{x}} = \mathbf{F}\mathbf{y} + \mathbf{b}$.

It turns out that the augmented representation of widely linear estimators often shows structural similarities to their linear counterparts. This is the main reason why we favor the augmented form over the real composite representation. To motivate this further, a demonstration is presented in form of the LMMSE estimator and the WLMMSE estimator in the sequel. These estimators are discussed in detail in Chapter 4, thus we only show the formal expressions for now. The LMMSE estimator is given by

$$\hat{\mathbf{x}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}^{-1}_{\mathbf{yy}}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{2.24}$$

Its widely linear counterpart, the WLMMSE estimator, is most compactly written in augmented form as [2,22]

$$\underline{\hat{\mathbf{x}}} = E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]). \tag{2.25}$$

Note the elegant representation and the close notational similarity to the linear estimator in (2.24).

2.2.4 Gaussian Random Vectors

To simplify notation we regard zero mean vectors in the following. The Gaussian PDF of the real composite $2N_{\mathbf{x}}$-dimensional vector $\mathbf{x}_{\mathbb{R}} = \begin{bmatrix} \mathbf{x}_{\mathrm{R}}^T & \mathbf{x}_{\mathrm{I}}^T \end{bmatrix}^T$ is [2,23]

$$p(\mathbf{x}_{\mathbb{R}}) = \frac{1}{(2\pi)^{\frac{2N_{\mathbf{x}}}{2}}\sqrt{\det \mathbf{C}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}}}} \exp\left\{-\frac{1}{2}\mathbf{x}_{\mathbb{R}}^T \mathbf{C}^{-1}_{\mathbf{x}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}} \mathbf{x}_{\mathbb{R}}\right\}. \tag{2.26}$$

Using $\mathbf{x}_\mathbb{R} = \frac{1}{2}\mathbf{T}_{N_\mathbf{x}}^H \underline{\mathbf{x}}$, $\mathbf{C}_{\mathbf{x}_\mathbb{R}\mathbf{x}_\mathbb{R}}^{-1} = \mathbf{T}_{N_\mathbf{x}}^H \underline{\mathbf{C}}_{\mathbf{xx}}^{-1} \mathbf{T}_{N_\mathbf{x}}$, and $\det \mathbf{C}_{\mathbf{x}_\mathbb{R}\mathbf{x}_\mathbb{R}} = 2^{-2N_\mathbf{x}} \det \underline{\mathbf{C}}_{\mathbf{xx}}$, we obtain the PDF of the complex-valued vector $\mathbf{x}$ as [24,25]

$$p(\mathbf{x}) = \frac{1}{\pi^{N_\mathbf{x}} \sqrt{\det \underline{\mathbf{C}}_{\mathbf{xx}}}} \exp\left\{ -\frac{1}{2} \underline{\mathbf{x}}^H \underline{\mathbf{C}}_{\mathbf{xx}}^{-1} \underline{\mathbf{x}} \right\}. \tag{2.27}$$

Algebraically, this PDF depends on $\underline{\mathbf{x}}$, and thus on $\mathbf{x}$ and $\mathbf{x}^*$, but is interpreted as the joint PDF of $\mathbf{x}_\mathrm{R}$ and $\mathbf{x}_\mathrm{I}$. It can be used for proper or improper $\mathbf{x}$. In this work, we call a complex vector $\mathbf{x}$ with the distribution in (2.27) *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 $\mathbf{x}$ as

$$p(\mathbf{x}) = \frac{1}{\pi^{N_\mathbf{x}} \det \mathbf{C}_{\mathbf{xx}}} \exp\left\{ -\mathbf{x}^H \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{x} \right\}. \tag{2.28}$$

If it holds that $E_\mathbf{x}[\mathbf{x}] = \mathbf{0}$ and $\mathbf{C}_{\mathbf{xx}} = \mathbf{I}$ we simply refer to (2.28) as *standard* proper Gaussian PDF.

2.3 Wirtinger Calculus

When deriving an estimator it is often necessary to find the minimum of a real-valued cost function. This can in many cases be done by setting the gradient of the cost function equal to zero and derive the corresponding estimator. Matters turn more complicated when the estimator, the quantity we are interested in, becomes complex-valued. Then, the question is how the real-valued cost function can be differentiated w.r.t. this complex-valued quantity. The key to tackle this problem was provided by Wirtinger in 1927 [4]. In the following, we shortly summarize the main aspects of Wirtinger calculus. This section is basically a shortened version of Appendix 2 in [2], which itself is based on [4,26,27].

A main result of classical[1] complex analysis is that a complex-valued function is *complex-differentiable* on its entire domain if and only if it is *holomorphic*. Since non-constant, real-valued functions defined on the complex domain cannot be holomorphic, their classical complex derivations do not exist.

However, there exists a way to overcome this problem. Let the real-valued function f be defined on $\mathbb{C}^n$. By considering the real and imaginary parts of the n complex variables as separate variables, then f is defined on $\mathbb{R}^{2n}$. If f is differentiable on $\mathbb{R}^{2n}$ it is termed *real-differentiable*. According to classical complex analysis, a real-differentiable function is also complex-differentiable if and only if the *Cauchy-Riemann equations* hold. For real-differentiable functions, Wilhelm Wirtinger showed a way to define a generalized complex derivative which can also be conducted in situations where the Cauchy-Riemann equations do not hold. This *generalized complex derivative* exists whenever f is real-differentiable.

[1] Note that in this section the term 'classical' is used to indicate complex analysis approaches for differentiation without considering Wirtinger calculus and it is unrelated to classical estimation theory.

In the following, the Wirtinger calculus is derived for the scalar case. An extension to the vector case is presented afterwards.

Scalar Case

Let f be a real-valued scalar function of the complex-valued scalar $x = x_\mathrm{R} + jx_\mathrm{I}$. We denote $\mathbf{x}_\mathbb{R} = \begin{bmatrix} x_\mathrm{R} & x_\mathrm{I} \end{bmatrix}^T$ and write $f(x) = f(\mathbf{x}_\mathbb{R}) = f(x_\mathrm{R}, x_\mathrm{I})$ for convenience. As preparation for the generalized complex differential operator, we consider the task of linearly approximating the function $f(\mathbf{x}_\mathbb{R})$ around $\mathbf{x}_{\mathbb{R},0} = \begin{bmatrix} x_{\mathrm{R},0} & x_{\mathrm{I},0} \end{bmatrix}^T$

$$f(\mathbf{x}_\mathbb{R}) \approx f(\mathbf{x}_{\mathbb{R},0}) + \nabla_{\mathbf{x}_\mathbb{R}} f(\mathbf{x}_{\mathbb{R},0})(\mathbf{x}_\mathbb{R} - \mathbf{x}_{\mathbb{R},0}). \tag{2.29}$$

$\nabla_{\mathbf{x}_\mathbb{R}}$ is the real differential operator defined as the row vector

$$\nabla_{\mathbf{x}_\mathbb{R}} f(\mathbf{x}_{\mathbb{R},0}) = \begin{bmatrix} \frac{\partial f}{\partial x_\mathrm{R}}(\mathbf{x}_{\mathbb{R},0}) & \frac{\partial f}{\partial x_\mathrm{I}}(\mathbf{x}_{\mathbb{R},0}) \end{bmatrix}. \tag{2.30}$$

We note that in engineering literature, the gradient of a function is in many cases defined to be a column vector. Whereas in mathematics literature a gradient is usually defined to be a row vector. In this work, we define the gradient of a scalar valued function to be a row vector as in (2.30) [2].

From $\mathbf{x}_\mathbb{R}$, the augmented vector $\underline{\mathbf{x}} = \begin{bmatrix} x & x^* \end{bmatrix}^T$ is obtained via

$$\underline{\mathbf{x}} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \mathbf{x}_\mathbb{R} = \mathbf{T}_1 \mathbf{x}_\mathbb{R}. \tag{2.31}$$

By utilizing $\mathbf{T}_1^H \mathbf{T}_1 = \mathbf{T}_1 \mathbf{T}_1^H = 2\mathbf{I}$, the right term in (2.29) reads as

$$\begin{aligned} \nabla_{\mathbf{x}_\mathbb{R}} f(\mathbf{x}_{\mathbb{R},0})(\mathbf{x}_\mathbb{R} - \mathbf{x}_{\mathbb{R},0}) &= \left(\frac{1}{2} \nabla_{\mathbf{x}_\mathbb{R}} f(\mathbf{x}_{\mathbb{R},0}) \mathbf{T}_1^H\right)(\mathbf{T}_1(\mathbf{x}_\mathbb{R} - \mathbf{x}_{\mathbb{R},0})) & (2.32) \\ &= \left(\frac{1}{2} \nabla_{\mathbf{x}_\mathbb{R}} f(\mathbf{x}_{\mathbb{R},0}) \mathbf{T}_1^H\right)(\underline{\mathbf{x}} - \underline{\mathbf{x}}_0). & (2.33) \end{aligned}$$

This result motivates the definition of the *complex gradient* as

$$\begin{aligned} \nabla_x f(x_0) &= \frac{1}{2} \nabla_{\mathbf{x}_\mathbb{R}} f(\mathbf{x}_{\mathbb{R},0}) \mathbf{T}_1^H & (2.34) \\ &= \begin{bmatrix} \frac{1}{2}\left(\frac{\partial f}{\partial x_\mathrm{R}} - j\frac{\partial f}{\partial x_\mathrm{I}}\right)(\mathbf{x}_{\mathbb{R},0}) & \frac{1}{2}\left(\frac{\partial f}{\partial x_\mathrm{R}} + j\frac{\partial f}{\partial x_\mathrm{I}}\right)(\mathbf{x}_{\mathbb{R},0}) \end{bmatrix}. & (2.35) \end{aligned}$$

Combining (2.33) and (2.35) reveals that the first operator in (2.35) is applied on x and the second operator is applied on x^*. This motivates the definition of the *generalized complex differential* operator as

$$\frac{\partial}{\partial x} = \frac{1}{2}\left(\frac{\partial}{\partial x_\mathrm{R}} - j\frac{\partial}{\partial x_\mathrm{I}}\right) \tag{2.36}$$

and the *conjugate generalized complex differential* operator as

$$\frac{\partial}{\partial x^*} = \frac{1}{2}\left(\frac{\partial}{\partial x_{\mathrm{R}}} + j\frac{\partial}{\partial x_{\mathrm{I}}}\right), \tag{2.37}$$

such that the gradient in (2.35) reads as

$$\nabla_x f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x}(x_0) & \frac{\partial f}{\partial x^*}(x_0) \end{bmatrix}. \tag{2.38}$$

The results in (2.36) and (2.37) are sometimes referred to as *Wirtinger derivative* and *conjugate Wirtinger derivative*, respectively. We use the plural form *Wirtinger derivatives* to account for both expressions. The following short example reveals an interesting effect when applying the Wirtinger derivatives.

Let $f(x)$ be given by $f(x) = |x|^2 = xx^*$. With $f(x) = f(x_{\mathrm{R}}, x_{\mathrm{I}}) = x_{\mathrm{R}}^2 + x_{\mathrm{I}}^2$ the Wirtinger derivatives follow to

$$\frac{\partial f(x)}{\partial x} = \frac{1}{2}\left(\frac{\partial}{\partial x_{\mathrm{R}}} - j\frac{\partial}{\partial x_{\mathrm{I}}}\right) f(x_{\mathrm{R}}, x_{\mathrm{I}}) = x_{\mathrm{R}} - jx_{\mathrm{I}} = x^* \tag{2.39}$$

$$\frac{\partial f(x)}{\partial x^*} = \frac{1}{2}\left(\frac{\partial}{\partial x_{\mathrm{R}}} + j\frac{\partial}{\partial x_{\mathrm{I}}}\right) f(x_{\mathrm{R}}, x_{\mathrm{I}}) = x_{\mathrm{R}} + jx_{\mathrm{I}} = x. \tag{2.40}$$

This result is interesting since the same expressions can be derived by formally treating x and x^* as two independent variables. This suggests to treat x^* as a constant when applying $\frac{\partial}{\partial x}$, and to treat x as a constant when applying $\frac{\partial}{\partial x^*}$. Following this guidance leads to

$$\frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x}(xx^*) = x^* \tag{2.41}$$

$$\frac{\partial f(x)}{\partial x^*} = \frac{\partial}{\partial x^*}(xx^*) = x, \tag{2.42}$$

which are equal to (2.39) and (2.40), respectively. Hence, the Wirtinger derivatives (for this example) can be derived by treating x and x^* as two independent variables. Ultimately, it can be shown that this statement holds in general.

The frequent task in this work is to find the minimum of a real-valued cost function $f(\mathbf{x}_{\mathbb{R}})$. This can be done by setting the gradient equal to zero, i.e.

$$\nabla_{\mathbf{x}_{\mathbb{R}}} f(\mathbf{x}_{\mathbb{R}}) = \begin{bmatrix} \frac{\partial f}{\partial x_{\mathrm{R}}}(\mathbf{x}_{\mathbb{R}}) & \frac{\partial f}{\partial x_{\mathrm{I}}}(\mathbf{x}_{\mathbb{R}}) \end{bmatrix} \overset{!}{=} \begin{bmatrix} 0 & 0 \end{bmatrix}. \tag{2.43}$$

This implies setting two equations equal to zero and solving for $\mathbf{x}_{\mathbb{R}}$. In many cases, simpler minimizations processes and more compact analytical expressions result by utilizing the Wirtinger derivatives. Considering the definition in (2.36), one can state that setting the Wirtinger derivative equal to zero

$$\frac{\partial f(x)}{\partial x} = 0 \tag{2.44}$$

is equivalent to (2.43) but only requires to solve a single complex-valued equation.

Although we focused on real-valued functions $f(x)$ to demonstrate the need of generalizing the complex differential operator, there is nothing in (2.29)–(2.42) that prevents applying this concept to complex-valued functions. Hence, the Wirtinger derivatives can be applied to complex-valued functions, too. It can easily be shown [2], that for a holomorphic function, the Wirtinger derivative is the standard complex derivative.

Vector Case

Till now only the scalar case was investigated. For the vector case considered in the following, we only show the main results and definitions here. Let $f : \mathbb{C}^{N_\mathbf{x}} \to \mathbb{C}$ be a complex-valued scalar function of the complex-valued vector $\mathbf{x} = \mathbf{x}_\mathrm{R} + j\mathbf{x}_\mathrm{I}$. Further, let $\mathbf{x}_\mathbb{R} = \begin{bmatrix}\mathbf{x}_\mathrm{R}^T & \mathbf{x}_\mathrm{I}^T\end{bmatrix}^T$. If $\mathrm{Re}\{f(\mathbf{x}_\mathbb{R})\}$ and $\mathrm{Im}\{f(\mathbf{x}_\mathbb{R})\}$ are both real-differentiable, the complex gradient is a $1 \times 2N_\mathbf{x}$ row vector given as

$$\nabla_\mathbf{x} f = \frac{\partial f}{\partial \underline{\mathbf{x}}} = \begin{bmatrix}\frac{\partial f}{\partial \mathbf{x}} & \frac{\partial f}{\partial \mathbf{x}^*}\end{bmatrix}, \tag{2.45}$$

where

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_{N_\mathbf{x}}}\end{bmatrix}, \tag{2.46}$$

$$\frac{\partial f}{\partial \mathbf{x}^*} = \begin{bmatrix}\frac{\partial f}{\partial x_1^*} & \frac{\partial f}{\partial x_2^*} & \cdots & \frac{\partial f}{\partial x_{N_\mathbf{x}}^*}\end{bmatrix}. \tag{2.47}$$

For $\mathbf{f}$ being a vector valued function $\mathbf{f} : \mathbb{C}^{N_\mathbf{x}} \to \mathbb{C}^m$, the complex Jacobian is defined as the $m \times 2N_\mathbf{x}$ matrix

$$\mathbf{J}_\mathbf{x} = \begin{bmatrix}\nabla_\mathbf{x} f_1 \\ \nabla_\mathbf{x} f_2 \\ \vdots \\ \nabla_\mathbf{x} f_m\end{bmatrix}. \tag{2.48}$$

In the following, we list some rules and special cases for Wirtinger derivatives. By doing so, we assume the arbitrary vectors $\mathbf{a}$ and $\mathbf{b}$ as well as the arbitrary matrix $\mathbf{A}$ are independent of $\mathbf{x}$ and $\mathbf{x}^*$. At first we consider the case $f : \mathbb{C}^{N_\mathbf{x}} \to \mathbb{C}$. Then, it holds that [28]

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{a}^H \mathbf{x} = \mathbf{a}^H \qquad \frac{\partial}{\partial \mathbf{x}^*} \mathbf{a}^H \mathbf{x} = \mathbf{0}^T, \tag{2.49}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^H \mathbf{a} = \mathbf{0}^T \qquad \frac{\partial}{\partial \mathbf{x}^*} \mathbf{x}^H \mathbf{a} = \mathbf{a}^T, \tag{2.50}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{a} = \mathbf{a}^T \qquad \frac{\partial}{\partial \mathbf{x}^*} \mathbf{x}^T \mathbf{a} = \mathbf{0}^T, \tag{2.51}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^H \mathbf{A} \mathbf{x} = \mathbf{x}^H \mathbf{A} \qquad \frac{\partial}{\partial \mathbf{x}^*} \mathbf{x}^H \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{A}^T, \tag{2.52}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \left(\mathbf{A} + \mathbf{A}^T\right) \qquad \frac{\partial}{\partial \mathbf{x}^*} \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{0}^T, \tag{2.53}$$

$$\frac{\partial}{\partial \mathbf{x}}\exp\left(-\frac{1}{2}\mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}\right) = -\frac{1}{2}\exp\left(-\frac{1}{2}\mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}\right)\mathbf{x}^H\mathbf{A}^{-1}, \tag{2.54}$$

$$\frac{\partial}{\partial \mathbf{x}}\ln\left(\mathbf{x}^H\mathbf{A}\mathbf{x}\right) = \left(\mathbf{x}^H\mathbf{A}\mathbf{x}\right)^{-1}\mathbf{x}^H\mathbf{A}. \tag{2.55}$$

For vector valued functions $\mathbf{f} : \mathbb{C}^{N_\mathbf{x}} \to \mathbb{C}^m$ it holds that

$$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{I} \qquad \frac{\partial \mathbf{x}}{\partial \mathbf{x}^*} = \mathbf{0}, \tag{2.56}$$

$$\frac{\partial \mathbf{x}^*}{\partial \mathbf{x}} = \mathbf{0} \qquad \frac{\partial \mathbf{x}^*}{\partial \mathbf{x}^*} = \mathbf{I}, \tag{2.57}$$

$$\frac{\partial \mathbf{f}^*}{\partial \mathbf{x}^*} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)^* \qquad \frac{\partial \mathbf{f}}{\partial \mathbf{x}^*} = \left(\frac{\partial \mathbf{f}^*}{\partial \mathbf{x}}\right)^*. \tag{2.58}$$

The chain rules for two real-differentiable vector valued functions $\mathbf{f}$ and $\mathbf{g}$ are given by

$$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{x}} + \frac{\partial \mathbf{f}}{\partial \mathbf{g}^*}\frac{\partial \mathbf{g}^*}{\partial \mathbf{x}}, \tag{2.59}$$

$$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{x}^*} = \frac{\partial \mathbf{f}}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial \mathbf{x}^*} + \frac{\partial \mathbf{f}}{\partial \mathbf{g}^*}\frac{\partial \mathbf{g}^*}{\partial \mathbf{x}^*}. \tag{2.60}$$

If $\mathbf{f}$ is real-valued

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}^*} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)^* \tag{2.61}$$

holds since $\mathbf{f}(\mathbf{x}) = \mathbf{f}^*(\mathbf{x})$. Furthermore, for a real-valued scalar function f the following three conditions are equivalent

$$\nabla_\mathbf{x} f = \mathbf{0}^T \iff \frac{\partial f}{\partial \mathbf{x}} = \mathbf{0}^T \iff \frac{\partial f}{\partial \mathbf{x}^*} = \mathbf{0}^T. \tag{2.62}$$

As it will turn out later on, this result is important for finding local extrema of real-valued cost functions defined on the complex domain. It shows that the same local extrema are obtained when taking the derivative w.r.t. $\mathbf{x}$ or w.r.t. $\mathbf{x}^*$.

2.4 Lagrange Multiplier Method for the Complex Case

Many of the estimators in this work are derived by minimizing a cost function subject to some constraints. For such constrained optimization problems, the *Lagrange multiplier method* can be used to find an optimal solution. This method is a well-known optimization method, and can be found in many standard textbooks such as [29,30]. We consider the Lagrange multiplier method to be known and focus on its complex extension [6] in the following.

The task of optimizing a real-valued cost function $J(\mathbf{x})$ subject to a constraint $\mathbf{c}(\mathbf{x}) = \mathbf{0}$ is formally written as

$$\mathbf{x}_{\text{opt}} = \arg\min_{\mathbf{x}} J(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{c}(\mathbf{x}) = \mathbf{0} \tag{2.63}$$

and the *Lagrangian cost function* for real-valued constraints becomes

$$\mathcal{L}(\mathbf{x}) = J(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{x}), \tag{2.64}$$

where (the real-valued) $\boldsymbol{\lambda}$ is called the *Lagrange multiplier vector*. While the cost function $J(\mathbf{x})$ is always real-valued in this work, the function $\mathbf{c}(\mathbf{x})$ is in general complex-valued. In fact, this results in twice as many real-valued constraints, specifically

$$\mathrm{Re}\{\mathbf{c}(\mathbf{x})\} = \mathbf{0} \quad \text{and} \quad \mathrm{Im}\{\mathbf{c}(\mathbf{x})\} = \mathbf{0}. \tag{2.65}$$

Consequently, the appropriate Lagrangian cost function for this case reads as

$$\mathcal{L}(\mathbf{x}) = J(\mathbf{x}) + \boldsymbol{\lambda}_R^T \mathrm{Re}\{\mathbf{c}(\mathbf{x})\} + \boldsymbol{\lambda}_I^T \mathrm{Im}\{\mathbf{c}(\mathbf{x})\}. \tag{2.66}$$

The complex-valued Lagrange multiplier vector is defined as $\boldsymbol{\lambda} = \boldsymbol{\lambda}_R + j\boldsymbol{\lambda}_I$. Therewith, (2.66) can be rewritten as

$$\begin{aligned}
\mathcal{L}(\mathbf{x}) &= J(\mathbf{x}) + \frac{1}{2}\boldsymbol{\lambda}_R^T(\mathbf{c}(\mathbf{x}) + \mathbf{c}^*(\mathbf{x})) + \frac{1}{2}\boldsymbol{\lambda}_I^T(\mathbf{c}(\mathbf{x}) - \mathbf{c}^*(\mathbf{x})) && (2.67)\\
&= J(\mathbf{x}) + \frac{1}{2}\big(\boldsymbol{\lambda}_R^T + \boldsymbol{\lambda}_I^T\big)\mathbf{c}(\mathbf{x}) + \frac{1}{2}\big(\boldsymbol{\lambda}_R^T - \boldsymbol{\lambda}_I^T\big)\mathbf{c}^*(\mathbf{x}) && (2.68)\\
&= J(\mathbf{x}) + \frac{1}{4}\big(\boldsymbol{\lambda}^T + \boldsymbol{\lambda}^H + \boldsymbol{\lambda}^T - \boldsymbol{\lambda}^H\big)\mathbf{c}(\mathbf{x}) + \frac{1}{4}\big(\boldsymbol{\lambda}^T + \boldsymbol{\lambda}^H - \boldsymbol{\lambda}^T + \boldsymbol{\lambda}^H\big)\mathbf{c}^*(\mathbf{x}) && (2.69)\\
&= J(\mathbf{x}) + \frac{1}{2}\boldsymbol{\lambda}^T\mathbf{c}(\mathbf{x}) + \frac{1}{2}\boldsymbol{\lambda}^H\mathbf{c}^*(\mathbf{x}) && (2.70)\\
&= J(\mathbf{x}) + \frac{1}{2}\boldsymbol{\lambda}^T\mathbf{c}(\mathbf{x}) + \frac{1}{2}\big(\boldsymbol{\lambda}^T\mathbf{c}(\mathbf{x})\big)^*. && (2.71)
\end{aligned}$$

This result shows that formally $\mathbf{c}(\mathbf{x})$ as well as its conjugate have to be inserted into the Lagrangian cost function. We demonstrate the effects of this result with two upcoming examples that will appear later in this work in similar forms (e.g. in Section 3.2.1 and Section 3.4.1).

2.4.1 Linear Constraints

We optimize the real-valued cost function $J(\mathbf{x}) = \mathbf{x}^H\mathbf{A}\mathbf{x}$ with a Hermitian and invertible matrix $\mathbf{A}$ subject to the complex-valued linear constraint $\mathbf{B}\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \in \mathbb{C}^{N_\mathbf{b}}$ and where $\mathbf{B} \in \mathbb{C}^{N_\mathbf{b} \times N_\mathbf{x}}$ is a complex-valued matrix with $N_\mathbf{b} < N_\mathbf{x}$ and full row rank. Then, the resulting optimization problem formally reads

$$\mathbf{x}_{\mathrm{opt}} = \arg\min_{\mathbf{x}} \mathbf{x}^H\mathbf{A}\mathbf{x} \quad \text{s.t.} \quad \mathbf{B}\mathbf{x} = \mathbf{b}. \tag{2.72}$$

The corresponding Lagrangian cost function is given by

$$\mathcal{L}(\mathbf{x}) = \mathbf{x}^H\mathbf{A}\mathbf{x} + \frac{1}{2}\boldsymbol{\lambda}^T(\mathbf{B}\mathbf{x} - \mathbf{b}) + \frac{1}{2}\boldsymbol{\lambda}^H(\mathbf{B}^*\mathbf{x}^* - \mathbf{b}^*). \tag{2.73}$$

Taking the derivative w.r.t. $\mathbf{x}$ using the Wirtinger derivative yields

$$\frac{\partial \mathcal{L}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{x}^H\mathbf{A} + \frac{1}{2}\boldsymbol{\lambda}^T\mathbf{B}. \tag{2.74}$$

Setting this result equal to zero allows to identify $\mathbf{x}_{\text{opt}}$ as

$$\mathbf{x}_{\text{opt}}^H = -\frac{1}{2}\boldsymbol{\lambda}^T\mathbf{B}\mathbf{A}^{-1}, \tag{2.75}$$

$$\mathbf{x}_{\text{opt}} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{B}^H\boldsymbol{\lambda}^*. \tag{2.76}$$

Inserting this result into the constraint $\mathbf{Bx} = \mathbf{b}$ results in

$$-\frac{1}{2}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\boldsymbol{\lambda}^* = \mathbf{b}. \tag{2.77}$$

Due to the full row rank assumption of $\mathbf{B}$, the expression $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H$ in (2.77) is invertible and allows

$$-\frac{1}{2}\boldsymbol{\lambda}^* = \left(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\right)^{-1}\mathbf{b}. \tag{2.78}$$

Finally, reinserting into (2.76) yields

$$\mathbf{x}_{\text{opt}} = \mathbf{A}^{-1}\mathbf{B}^H\left(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\right)^{-1}\mathbf{b}. \tag{2.79}$$

Note that during the entire derivation the conjugate complex constraint in the Lagrangian cost function in (2.73) does not play a role. The reason for that is that this term does not depend on $\mathbf{x}$ but only on $\mathbf{x}^*$, making its derivative w.r.t. $\mathbf{x}$ vanishing. This conjugate complex constraint is required when taking the derivative w.r.t. $\mathbf{x}^*$. That is allowed since the Lagrangian cost function is real-valued, which allows to apply the rule in (2.62). Of course this would lead to the same result as in (2.79). Note further that the entire derivation could have been performed without the terms $\frac{1}{2}$ without changing the result. In other words, one may define a new variable $\boldsymbol{\lambda}' = \frac{1}{2}\boldsymbol{\lambda}$ to simplify the expressions. This will be done for the remainder of this work.

2.4.2 A Special Case of Widely Linear Constraints

We now replace our linear constraint $\mathbf{Bx} = \mathbf{b}$ by $\mathbf{Bx} + \mathbf{B}^*\mathbf{x}^* = \mathbf{b}$, where again $\mathbf{B} \in \mathbb{C}^{N_{\mathbf{b}} \times N_{\mathbf{x}}}$ fulfills $N_{\mathbf{b}} < N_{\mathbf{x}}$ and has full row rank. This constraint is frequently used in this work and it results in the optimization problem

$$\mathbf{x}_{\text{opt}} = \arg\min_{\mathbf{x}} \mathbf{x}^H\mathbf{A}\mathbf{x} \quad \text{s.t.} \quad \mathbf{Bx} + \mathbf{B}^*\mathbf{x}^* = \mathbf{b}. \tag{2.80}$$

The left hand side of the constraint is always real-valued, which enforces $\mathbf{b}$ to be real-valued, too. Thus, the Lagrangian cost function becomes

$$\mathcal{L}(\mathbf{x}) = \mathbf{x}^H\mathbf{A}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{Bx} + \mathbf{B}^*\mathbf{x}^* - \mathbf{b}) \tag{2.81}$$

with real $\boldsymbol{\lambda}$ since the introduction of an additional conjugate complex constraint as in (2.71) is obsolete. Taking the Wirtinger derivative of (2.81) w.r.t. $\mathbf{x}$ yields

$$\frac{\partial \mathcal{L}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{x}^H\mathbf{A} + \boldsymbol{\lambda}^T\mathbf{B}. \tag{2.82}$$

Setting this result equal to zero allows to identify

$$\mathbf{x}_{\text{opt}} = -\mathbf{A}^{-1}\mathbf{B}^{H}\boldsymbol{\lambda}. \tag{2.83}$$

Inserting this result into the constraint results in

$$-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{H}\boldsymbol{\lambda} - \mathbf{B}^{*}(\mathbf{A}^{*})^{-1}\mathbf{B}^{T}\boldsymbol{\lambda} = \mathbf{b}. \tag{2.84}$$

Assuming invertability of $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{H} + \mathbf{B}^{*}(\mathbf{A}^{*})^{-1}\mathbf{B}^{T}$, $\boldsymbol{\lambda}$ can be identified as

$$-\boldsymbol{\lambda} = \left(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{H} + \mathbf{B}^{*}(\mathbf{A}^{*})^{-1}\mathbf{B}^{T}\right)^{-1}\mathbf{b}. \tag{2.85}$$

This result reinserted into (2.83) yields

$$\mathbf{x}_{\text{opt}} = \mathbf{A}^{-1}\mathbf{B}^{H}\left(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{H} + \mathbf{B}^{*}(\mathbf{A}^{*})^{-1}\mathbf{B}^{T}\right)^{-1}\mathbf{b}. \tag{2.86}$$

3

Knowledge-Aided Concepts in Classical Estimation

This chapter focuses on aspects in classical estimation with an underlying linear model, where the unknown parameter vector is considered to be deterministic. We start by recapitulating the well-known LS and weighted least squares (WLS) estimators. Both estimators are purely deterministic and do not incorporate any statistics. This is in contrast to the BLUE and the BWLUE, which incorporate (augmented) second order statistics of the noise. Subsequently, we recapitulate commutation properties of the BLUE over linear transformations. Standard literature [1] states that the BLUE commutes over rectangular transformation matrices $\mathbf{B} \in \mathbb{C}^{m \times n}$ with $m \leq n$. For these cases, a parameter vector with dimension n is estimated and then transformed into another parameter vector with dimension $m \leq n$. The investigations start with a related problem where it is known that the parameter vector of length n lies in a linear subspace of $\mathbb{C}^n$ with dimension m and $m < n$. Let this subspace be spanned by the columns of the full column rank matrix $\mathbf{B}' \in \mathbb{C}^{n \times m}$. A straight forward approach to incorporate this additional model knowledge is to transform the estimated vector with length m into an estimate of the parameter vector with length n using $\mathbf{B}'$. An open question in standard literature is whether the resulting estimate of the parameter vector is the BLUE or not. In this work, it will be proven that this approach in fact produces the BLUE. For the LS estimator on the other hand it is shown that a constrained LS estimator [1] is able to incorporate the knowledge that the parameter vector lies in a linear subspace of $\mathbb{C}^n$. The linear constraints required for applying the constrained LS estimator are derived.

The contributions in this chapter continue with investigations of other possible sources of additional model knowledge. For example sometimes it is known that the parameter vector fulfills additional linear constraints. An example is the case where the parameter vector describes the impulse response of a linear system that is unable to transmit direct current (DC) signals. Then, the sum of the taps of the impulse response must be zero, which can be reformulated as a linear constraint. In literature the constrained LS estimator is available as a standard estimator for this case, however, to the best of our knowledge constrained versions of the BLUE and the BWLUE have not been published so far. In this work, this gap is closed by introducing the constrained BLUE and the constrained BWLUE.

The next investigations consider the task of estimating a real-valued parameter vector, while the measurements shall remain complex-valued. A prominent example is estimating a real-valued impulse response of a system based on complex-valued measurements of its frequency response [31]. Applying the ordinary BLUE for this task in general results in complex-valued estimates, producing a systematic error. We propose several classical estimators that avoid this systematic error by producing real-valued estimates only, increasing the estimation accuracy compared to standard estimators as well as intuitive estimators.

Finally, we consider the case where the measurement matrix is not completely known but rather disturbed by additive noise [7–12]. Typical practical applications are problems, for which the measurement matrix is for example a convolution matrix that is constructed based on an imperfectly measured or estimated impulse response. We propose an iterative algorithm that incorporates the error variances into the estimation process. It is demonstrated that this algorithm outperforms competing algorithms significantly.

3.1 State-of-the-Art

Consider the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ is a complex-valued parameter vector, $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ is a complex-valued measurement vector, $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a complex-valued measurement matrix with full column rank and[2] $N_\mathbf{x} < N_\mathbf{y}$, and $\mathbf{n} \in \mathbb{C}^{N_\mathbf{y}}$ is a complex-valued random proper noise vector with zero mean. Later on, we will dismiss the proper noise assumption.

3.1.1 Classical Estimators

In the following, we recapitulate the well-known classical estimator for the model in (3.1).

LS Estimator

In LS estimation the sum of the absolute squared differences between the elements of the actual measurements $\mathbf{y}$ and the elements of the *assumed signal* or *noiseless data* $\mathbf{s} = \mathbf{H}\mathbf{x} \in \mathbb{C}^{N_\mathbf{y}}$ is minimized [1]. Let y_i and s_i be the i^{th} elements of $\mathbf{y}$ and $\mathbf{s}$, respectively.

[2]We note that the investigations could also be done with $N_\mathbf{x} \leq N_\mathbf{y}$, but when $N_\mathbf{x} = N_\mathbf{y}$ the BLUE reduces to the simple inverse of the measurement matrix.

Then, the cost function follows as

$$J(\mathbf{x}) = \sum_{i=1}^{N_\mathbf{y}} |y_i - s_i|^2 \tag{3.2}$$

$$= (\mathbf{y} - \mathbf{H}\mathbf{x})^H (\mathbf{y} - \mathbf{H}\mathbf{x}). \tag{3.3}$$

The vector $\mathbf{x}$ that minimizes this cost function is denoted as the LS estimator determined as

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \arg \min_{\mathbf{x}} \; (\mathbf{y} - \mathbf{H}\mathbf{x})^H (\mathbf{y} - \mathbf{H}\mathbf{x}). \tag{3.4}$$

The solution is given by [1]

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{y} \tag{3.5}$$

$$= \mathbf{E}_{\mathrm{LS}} \mathbf{y}, \tag{3.6}$$

where $\mathbf{H}^H\mathbf{H}$ is invertible since $\mathbf{H}$ is assumed to have full column rank. This estimator is linear in the measurements $\mathbf{y}$ and the linear operator $\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ is termed the estimator matrix. This estimator matrix is shortly written as $\mathbf{E}_{\mathrm{LS}}$ such that $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{E}_{\mathrm{LS}}\mathbf{y}$. Note that the LS estimator in (3.5) does not require any statistical knowledge about the measurements. However, first and second order statistics of the noise are required when inspecting the usual performance measures.

The performance of an estimator is usually measured in terms of the MSEs between the elements of the estimated and the true parameter vector. Deriving the MSEs requires averaging over the PDF of the measurement vector $\mathbf{y}$. According to (3.1), the PDF of $\mathbf{y}$ corresponds to the PDF of $\mathbf{n}$ shifted such that its mean is $\mathbf{H}\mathbf{x}$. Hence, statistics about $\mathbf{n}$ are required. Besides the zero mean assumption already introduced, we further assume the noise covariance matrix $\mathbf{C}_{\mathbf{nn}} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{y}}$ to be known. With this knowledge, the MSEs of the elements of $\hat{\mathbf{x}}$ can be derived analytically. Let $\hat{x}_i$ and x_i be the i^{th} elements of $\hat{\mathbf{x}}$ and $\mathbf{x}$, respectively. Then, the MSE of $\hat{x}_i$ can be separated into the sum of its variance $\mathrm{var}(\hat{x}_i)$ and absolute squared bias $|\mathrm{b}(\hat{x}_i)|^2 = |E_\mathbf{y}[\hat{x}_i] - x_i|^2$ according to [1]

$$\mathrm{mse}(\hat{x}_i) = \mathrm{var}(\hat{x}_i) + |\mathrm{b}(\hat{x}_i)|^2. \tag{3.7}$$

Hence, for unbiased estimators $\mathrm{b}(\hat{x}_i) = 0$, the MSEs $\mathrm{mse}(\hat{x}_i)$ correspond to the variances $\mathrm{var}(\hat{x}_i)$. It turns out that the LS estimator for the assumed model is unbiased, which corresponds to

$$E_\mathbf{y}[\hat{\mathbf{x}}_{\mathrm{LS}}] = \mathbf{x}. \tag{3.8}$$

As a consequence, the covariance matrix of the vector estimator $\hat{\mathbf{x}}_{\mathrm{LS}}$ is given by [1]

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}} = E_\mathbf{y}\left[(\hat{\mathbf{x}}_{\mathrm{LS}} - E_\mathbf{y}[\hat{\mathbf{x}}_{\mathrm{LS}}])(\hat{\mathbf{x}}_{\mathrm{LS}} - E_\mathbf{y}[\hat{\mathbf{x}}_{\mathrm{LS}}])^H\right] \tag{3.9}$$

$$= \mathbf{E}_{\mathrm{LS}} \mathbf{C}_{\mathbf{nn}} \mathbf{E}_{\mathrm{LS}}^H. \tag{3.10}$$

The variances of the individual elements of $\hat{\mathbf{x}}_{\mathrm{LS}}$, which correspond to the MSEs, can be found on the main diagonal of the covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}}$.

Weighted LS Estimator

The cost function in (3.2) treats every measurement equally. If it is known that some measurements are more reliable than others it makes sense to modify this cost function by giving those reliable measurements more weight. This can be done by assigning individual weights w_i to the measurements y_i. These weights are assumed to be real-valued, positive, non-zero and finite. The diagonal matrix $\mathbf{W}$ with the weights w_i assembled on the main diagonal is denoted as weighting matrix. This notation allows to modify the cost function in (3.2) as

$$J(\mathbf{x}) = \sum_{i=1}^{N_{\mathbf{y}}} w_i |y_i - s_i|^2 \tag{3.11}$$

$$= (\mathbf{y} - \mathbf{Hx})^H \mathbf{W} (\mathbf{y} - \mathbf{Hx}). \tag{3.12}$$

The WLS estimator is given by the vector $\mathbf{x}$ that minimizes this cost function

$$\hat{\mathbf{x}}_{\text{WLS}} = \arg \min_{\mathbf{x}} \; (\mathbf{y} - \mathbf{Hx})^H \mathbf{W} (\mathbf{y} - \mathbf{Hx}) \tag{3.13}$$

and follows as [1]

$$\hat{\mathbf{x}}_{\text{WLS}} = \left(\mathbf{H}^H \mathbf{W} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{W} \mathbf{y} \tag{3.14}$$

$$= \mathbf{E}_{\text{WLS}} \mathbf{y}. \tag{3.15}$$

The WLS estimator for the assumed model is also unbiased and its covariance matrix is

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\text{WLS}} = \mathbf{E}_{\text{WLS}} \mathbf{C}_{\mathbf{nn}} \mathbf{E}_{\text{WLS}}^H. \tag{3.16}$$

The product $\mathbf{H}^H \mathbf{W} \mathbf{H}$ is always invertible due to the full column rank assumption of $\mathbf{H}$ and due to the assumptions we have made on the weights w_i.

Constrained LS Estimator

In many applications it is known that the parameter vector fulfills some constraints, e.g., the linear constraints

$$\mathbf{Ax} = \mathbf{b}, \tag{3.17}$$

where $\mathbf{A} \in \mathbb{C}^{N_{\mathbf{b}} \times N_{\mathbf{x}}}$, $\mathbf{b} \in \mathbb{C}^{N_{\mathbf{b}}}$, $N_{\mathbf{b}} < N_{\mathbf{x}}$. Since the parameter vector is assumed to fulfill (3.17), we seek for an estimator whose estimates fulfill

$$\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{3.18}$$

A modification of the LS estimator that fulfills (3.18) can be found in [1] and its derivation will be repeated in the following.

The ordinary LS estimator is derived by minimizing the LS cost function in (3.3). This cost function is now minimized subject to the additional constraint. The optimization task is formally given by

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x}} \; (\mathbf{y} - \mathbf{H}\mathbf{x})^H(\mathbf{y} - \mathbf{H}\mathbf{x}) \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{3.19}$$

In addition to the assumptions already mentioned, we constrain the matrix $\mathbf{A}$ to have linearly independent rows such that the matrix $\mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H$ is invertible. From (3.19), the Lagrangian cost function follows as

$$\begin{aligned}
\mathcal{L}(\mathbf{x}) &= (\mathbf{y} - \mathbf{H}\mathbf{x})^H(\mathbf{y} - \mathbf{H}\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\lambda}^H(\mathbf{A}^*\mathbf{x}^* - \mathbf{b}^*) &(3.20)\\
&= \mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{H}\mathbf{x} - \mathbf{x}^H\mathbf{H}^H\mathbf{y} + \mathbf{x}^H\mathbf{H}^H\mathbf{H}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\lambda}^H(\mathbf{A}^*\mathbf{x}^* - \mathbf{b}^*). &(3.21)
\end{aligned}$$

Then, the Wirtinger derivative of (3.21) w.r.t. $\mathbf{x}$ is given by

$$\frac{\partial \mathcal{L}(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{y}^H\mathbf{H} + \mathbf{x}^H\mathbf{H}^H\mathbf{H} + \boldsymbol{\lambda}^T\mathbf{A}. \tag{3.22}$$

Setting this result equal to zero yields the estimator

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\boldsymbol{\lambda}^*. \tag{3.23}$$

Inserting this result into $\mathbf{A}\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{b}$ allows to identify $\boldsymbol{\lambda}$ as

$$\begin{aligned}
\mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} - \underbrace{\mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H}_{\mathbf{K}}\boldsymbol{\lambda}^* &= \mathbf{b} &(3.24)\\
-\mathbf{K}\boldsymbol{\lambda}^* &= \mathbf{b} - \mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} &(3.25)\\
-\boldsymbol{\lambda}^* &= \mathbf{K}^{-1}\mathbf{b} - \mathbf{K}^{-1}\mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y}, &(3.26)
\end{aligned}$$

where

$$\mathbf{K} = \mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H. \tag{3.27}$$

A reinsertion of (3.26) into (3.23) yields

$$\begin{aligned}
\hat{\mathbf{x}}_{\mathrm{LS}} &= \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} + \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{b}\\
&\quad - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} &(3.28)\\
&= \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} + \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{b}, &(3.29)
\end{aligned}$$

which represents the final result for the constrained LS estimator.

Deriving the covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}}$ requires noise statistics in form of $\mathbf{C}_{\mathbf{nn}}$ to be known. Note that (3.29) fulfills $E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}] = \mathbf{x}$. With that, an intermediate result can be derived as

$$\hat{\mathbf{x}}_{\mathrm{LS}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}] = \hat{\mathbf{x}}_{\mathrm{LS}} - \mathbf{x} \tag{3.30}$$

$$= \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} + \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{b} - \mathbf{x} \tag{3.31}$$

$$= \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{H}\mathbf{x} + \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{n} + \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{b} - \mathbf{x} \tag{3.32}$$

$$= \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{n}. \tag{3.33}$$

Now, $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}}$ becomes

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}} = E_{\mathbf{y}}\Big[(\hat{\mathbf{x}}_{\mathrm{LS}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}])(\hat{\mathbf{x}}_{\mathrm{LS}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}])^H\Big] \tag{3.34}$$

$$= \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}} \times \mathbf{H}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\Big(\mathbf{I} - \mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\Big). \tag{3.35}$$

BLUE

For deriving the performance measures of the LS estimator the noise covariance matrix $\mathbf{C}_{\mathbf{nn}}$ was assumed to be known. The estimator discussed in the following is able to incorporate this covariance matrix already into the estimation process. The goal is to find a linear estimator $\hat{\mathbf{x}} = \mathbf{E}\mathbf{y}$ that is unbiased and that minimizes the variances of the estimates $\hat{x}_i$, where $\hat{x}_i$ is the i^{th} element of $\hat{\mathbf{x}}$. In fact, this approach will lead to the famous BLUE. In that sense, the BLUE is best at minimizing the variance of the estimates among all linear and unbiased estimators. We assume that the linear model in (3.1) holds and we focus on $\hat{x}_i$ first. This scalar $\hat{x}_i$ is connected with the measurements via $\hat{x}_i = \mathbf{e}_i^H\mathbf{y}$, where $\mathbf{e}_i^H$ is the i^{th} row of the estimator matrix $\mathbf{E}$. We now analyze the unbiased condition on $\hat{x}_i$. For the model in (3.1) we obtain

$$E_{\mathbf{y}}[\hat{x}_i] = E_{\mathbf{y}}\big[\mathbf{e}_i^H\mathbf{y}\big] \tag{3.36}$$

$$= E_{\mathbf{n}}\big[\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n}\big] \tag{3.37}$$

$$= \mathbf{e}_i^H\mathbf{H}\mathbf{x} \overset{!}{=} x_i. \tag{3.38}$$

To fulfill this for every $\mathbf{x}$ the condition $\mathbf{e}_i^H\mathbf{H} = \mathbf{u}_i^T$ must hold, where $\mathbf{u}_i^T$ is a row vector of size $1 \times N_{\mathbf{x}}$ with a '1' at its i^{th} position, and all zeros elsewhere. The cost function, which is the variance of $\hat{x}_i$, becomes

$$J(\mathbf{e}_i) = E_{\mathbf{y}}\left[(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])^H\right] \tag{3.39}$$
$$= E_{\mathbf{y}}\left[(\hat{x}_i - x_i)(\hat{x}_i - x_i)^H\right] \tag{3.40}$$
$$= E_{\mathbf{y}}\left[\left(\mathbf{e}_i^H\mathbf{y} - x_i\right)\left(\mathbf{e}_i^H\mathbf{y} - x_i\right)^H\right] \tag{3.41}$$
$$= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} - x_i\right)\left(\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} - x_i\right)^H\right] \tag{3.42}$$
$$= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H\mathbf{n}\right)\left(\mathbf{e}_i^H\mathbf{n}\right)^H\right] \tag{3.43}$$
$$= \mathbf{e}_i^H\mathbf{C}_{\mathbf{nn}}\mathbf{e}_i. \tag{3.44}$$

The vector $\mathbf{e}_i$ that minimizes this cost function and that produces unbiased estimates is the solution of the constrained optimization problem

$$\mathbf{e}_{\mathrm{B},i} = \arg\min_{\mathbf{e}_i}\ \mathbf{e}_i^H\mathbf{C}_{\mathbf{nn}}\mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H\mathbf{H} = \mathbf{u}_i^T, \tag{3.45}$$

where the index B indicates the BLUE. Solving this constrained optimization problem using the Lagrange multiplier method described in Section 2.4 leads to the BLUE for x_i according to

$$\hat{x}_{\mathrm{B},i} = \mathbf{e}_{\mathrm{B},i}^H\mathbf{y} \tag{3.46}$$
$$= \mathbf{u}_i^T\left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}. \tag{3.47}$$

Since $\mathbf{u}_i^T$ is the only term that depends on the index i, the vector estimator immediately follows as

$$\hat{\mathbf{x}}_{\mathrm{B}} = \left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}, \tag{3.48}$$

which represents the final expression for the BLUE. The covariance matrix of the estimates is given by

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{B}} = \mathbf{E}_{\mathrm{B}}\mathbf{C}_{\mathbf{nn}}\mathbf{E}_{\mathrm{B}}^H \tag{3.49}$$
$$= \left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}, \tag{3.50}$$

where $\mathbf{E}_{\mathrm{B}} = \left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}$. Note the interesting similarities between the BLUE and the WLS estimator in (3.14). Also note that the estimator matrix of the BLUE fulfills $\mathbf{E}_{\mathrm{B}}\mathbf{H} = \mathbf{I}$.

The BLUE corresponds to the minimum variance unbiased (MVU) estimator for the linear model in (3.1) in the case of Gaussian distributed noise [1].

BWLUE

We will now dismiss the proper noise assumption and consider improper noise statistics instead. For this case, the LS and the WLS estimators do not change since both estimators do not incorporate the noise statistics at all. It can be shown that also the expressions for the corresponding covariance matrices in (3.10) and (3.16) do not change.

In Section 2.2, we stated that widely linear estimators can incorporate improper statistics. The BWLUE, which incorporates improper noise statistics, is the widely linear extension of the BLUE. Its derivation is summarized in the following. Consider the general widely linear estimator in (2.22) and its augmented notation in (2.23). Let $\mathbf{b}$ be the zero vector, and let $\mathbf{f}_i^H$, $\mathbf{g}_i^H$ and $\mathbf{e}_i^H$ denote the i^{th} rows of $\mathbf{F}$, $\mathbf{G}$ and $\underline{\mathbf{E}}$, respectively. Then, $\hat{x}_i$ is given by

$$\hat{x}_i = \begin{bmatrix} \mathbf{f}_i^H & \mathbf{g}_i^H \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} = \mathbf{e}_i^H \underline{\mathbf{y}}, \tag{3.51}$$

with $\mathbf{e}_i^H = \begin{bmatrix} \mathbf{f}_i^H & \mathbf{g}_i^H \end{bmatrix}$. For $\hat{x}_i$ to be unbiased it must hold that

$$E_{\mathbf{y}}[\hat{x}_i] = E_{\mathbf{n}}[\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H \underline{\mathbf{n}}] = \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} \stackrel{!}{=} x_i. \tag{3.52}$$

Hence, unbiasedness is ensured for every $\mathbf{x}$ if

$$\mathbf{e}_i^H \underline{\mathbf{H}} = \mathbf{u}_i^T, \tag{3.53}$$

where $\mathbf{u}_i^T$ is a row vector of size $1 \times 2N_{\mathbf{x}}$ with a '1' at its i^{th} position, and all zeros elsewhere.

The variance of $\hat{x}_i$ serves as the cost function that needs to be minimized and follows as [2]

$$\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y}}\left[(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])^H\right] && (3.54)\\
&= E_{\mathbf{y}}\left[(\hat{x}_i - x_i)(\hat{x}_i - x_i)^H\right] && (3.55)\\
&= E_{\mathbf{y}}\left[\left(\mathbf{e}_i^H \underline{\mathbf{y}} - x_i\right)\left(\mathbf{e}_i^H \underline{\mathbf{y}} - x_i\right)^H\right] && (3.56)\\
&= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H \underline{\mathbf{n}} - x_i\right)\left(\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H \underline{\mathbf{n}} - x_i\right)^H\right] && (3.57)\\
&= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H \underline{\mathbf{n}}\right)\left(\mathbf{e}_i^H \underline{\mathbf{n}}\right)^H\right] && (3.58)\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i. && (3.59)
\end{aligned}$$

The vector $\mathbf{e}_i$ that minimizes this cost function and that produces unbiased estimates is the solution of the constrained optimization problem

$$\mathbf{e}_{\mathrm{BW},i} = \arg \min_{\mathbf{e}_i} \; \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{H}} = \mathbf{u}_i^T, \tag{3.60}$$

where the index BW indicates the BWLUE. The optimization problem in (3.60) can be solved utilizing the Lagrange multiplier method described in Section 2.4. The solution directly leads to the BWLUE for x_i according to

$$\begin{aligned}
\hat{x}_{\mathrm{B},i} &= \mathbf{e}_{\mathrm{BW},i}^H \underline{\mathbf{y}} && (3.61)\\
&= \mathbf{u}_i^T \left(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\right)^{-1} \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{y}}. && (3.62)
\end{aligned}$$

Since $\mathbf{u}_i^T$ is the only term that depends on the index i, the vector BWLUE immediately follows as

$$\underline{\hat{\mathbf{x}}}_{\mathrm{B}} = \left(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\right)^{-1} \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{y}}. \tag{3.63}$$

It is important to note that this estimator is widely linear in $\mathbf{y}$ and incorporates the augmented covariance matrix of the noise. The estimators augmented covariance matrix is given by

$$\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{B}} = \underline{\mathbf{E}}_{\mathrm{BW}} \underline{\mathbf{C}}_{\mathbf{nn}} \underline{\mathbf{E}}_{\mathrm{BW}}^H \tag{3.64}$$

$$= \left(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\right)^{-1}, \tag{3.65}$$

where $\underline{\mathbf{E}}_{\mathrm{BW}} = \left(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\right)^{-1} \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}$. Note that the estimator matrix of the BWLUE fulfills $\underline{\mathbf{E}}_{\mathrm{BW}} \underline{\mathbf{H}} = \mathbf{I}$.

3.1.2 Commutation Analysis

In this section, we discuss cases for which the BLUE commutes over linear (actually affine) transformations. The arising results are then compactly extended to the BWLUE and to the LS estimator.

Commutation Analysis for the BLUE

In addition to the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.66}$$

we assume a new vector $\boldsymbol{\alpha}$ is connected with the parameter vector according to

$$\boldsymbol{\alpha} = \mathbf{B}\mathbf{x} + \mathbf{c} \tag{3.67}$$

with $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$ and $\mathbf{c} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$. We seek for the BLUE for $\boldsymbol{\alpha}$. In standard literature such as [1], it is stated that the BLUE for the linear model commutes over linear (actually affine) transformations as in (3.67) if either

1. $\mathbf{B} \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$ is an invertible matrix ($N_{\boldsymbol{\alpha}} = N_{\mathbf{x}}$) (Problem 6.12 in [1]), or

2. $\mathbf{B} \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$ with $N_{\boldsymbol{\alpha}} < N_{\mathbf{x}}$ and full row rank (Problem 4.12 in [1]).

If one of these two cases is fulfilled, and if the true parameter vector $\mathbf{x}$ is linearly transformed via (3.67), then the BLUE for the new vector $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}_{\mathrm{B}} = \mathbf{B}\hat{\mathbf{x}}_{\mathrm{B}} + \mathbf{c}. \tag{3.68}$$

In the following, we repeat the proof of both cases.

For the first case, reformulating (3.67) yields

$$\mathbf{x} = \mathbf{B}^{-1}(\boldsymbol{\alpha} - \mathbf{c}). \tag{3.69}$$

This expression inserted into the linear model in (3.1) produces

$$\mathbf{y} = \mathbf{H}\mathbf{B}^{-1}\boldsymbol{\alpha} - \mathbf{H}\mathbf{B}^{-1}\mathbf{c} + \mathbf{n} \tag{3.70}$$

$$\underbrace{\mathbf{y} + \mathbf{H}\mathbf{B}^{-1}\mathbf{c}}_{\widetilde{\mathbf{y}}} = \underbrace{\mathbf{H}\mathbf{B}^{-1}}_{\widetilde{\mathbf{H}}}\boldsymbol{\alpha} + \mathbf{n} \tag{3.71}$$

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{H}}\boldsymbol{\alpha} + \mathbf{n}. \tag{3.72}$$

For this modified linear model, the BLUE for $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}_{\mathrm{B}} = \left(\widetilde{\mathbf{H}}^H\mathbf{C}_{\mathbf{nn}}^{-1}\widetilde{\mathbf{H}}\right)^{-1}\widetilde{\mathbf{H}}^H\mathbf{C}_{\mathbf{nn}}^{-1}\widetilde{\mathbf{y}} \tag{3.73}$$

$$= \left(\left(\mathbf{B}^{-1}\right)^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\mathbf{B}^{-1}\right)^{-1}\left(\mathbf{B}^{-1}\right)^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\left(\mathbf{y} + \mathbf{H}\mathbf{B}^{-1}\mathbf{c}\right) \tag{3.74}$$

$$= \mathbf{B}\left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\ \right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\left(\mathbf{y} + \mathbf{H}\mathbf{B}^{-1}\mathbf{c}\right) \tag{3.75}$$

$$= \mathbf{B}\hat{\mathbf{x}}_{\mathrm{B}} + \mathbf{B}\mathbf{B}^{-1}\mathbf{c} \tag{3.76}$$

$$= \mathbf{B}\hat{\mathbf{x}}_{\mathrm{B}} + \mathbf{c}. \tag{3.77}$$

□

The second case can be proven as follows. Let the i^{th} elements of $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$ be denoted as α_i and $\hat{\alpha}_i$, respectively. Furthermore, the i^{th} row of the matrix $\mathbf{B}$ is denoted as $\mathbf{b}_i^H$ and the i^{th} element of $\mathbf{c}$ is denoted as c_i. We seek for an affine estimator of the form $\hat{\boldsymbol{\alpha}} = \mathbf{E}\mathbf{y} + \mathbf{d}$. Hence, the scalar $\hat{\alpha}_i$ is connected with the measurements via $\hat{\alpha}_i = \mathbf{e}_i^H\mathbf{y} + d_i$, where $\mathbf{e}_i^H$ is the i^{th} row of the estimator matrix $\mathbf{E}$ and where d_i is the i^{th} element of $\mathbf{d}$. Combining the linear model in (3.1) with (3.67) leads to

$$E_{\mathbf{y}}[\hat{\alpha}_i] = E_{\mathbf{y}}\left[\mathbf{e}_i^H\mathbf{y} + d_i\right] \tag{3.78}$$

$$= E_{\mathbf{n}}\left[\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} + d_i\right] \tag{3.79}$$

$$= \mathbf{e}_i^H\mathbf{H}\mathbf{x} + d_i \stackrel{!}{=} \alpha_i, \tag{3.80}$$

or

$$\mathbf{e}_i^H\mathbf{H}\mathbf{x} + d_i \stackrel{!}{=} \mathbf{b}_i^H\mathbf{x} + c_i. \tag{3.81}$$

To fulfill this for every $\mathbf{x}$ the conditions $\mathbf{e}_i^H\mathbf{H} = \mathbf{b}_i^H$ and $d_i = c_i$ must hold. The cost function, which is the variance of $\hat{\alpha}_i$, follows as

$$J(\mathbf{e}_i) = E_{\mathbf{y}}\left[(\hat{\alpha}_i - E_{\mathbf{y}}[\hat{\alpha}_i])(\hat{\alpha}_i - E_{\mathbf{y}}[\hat{\alpha}_i])^H\right] \tag{3.82}$$

$$= E_{\mathbf{y}}\left[\left(\mathbf{e}_i^H\mathbf{y} + d_i - E_{\mathbf{y}}\left[\mathbf{e}_i^H\mathbf{y} + d_i\right]\right)\left(\mathbf{e}_i^H\mathbf{y} + d_i - E_{\mathbf{y}}\left[\mathbf{e}_i^H\mathbf{y} + d_i\right]\right)^H\right] \tag{3.83}$$

$$= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} + d_i - \mathbf{e}_i^H\mathbf{H}\mathbf{x} - d_i\right)\left(\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} + d_i - \mathbf{e}_i^H\mathbf{H}\mathbf{x} - d_i\right)^H\right] \tag{3.84}$$

$$= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H\mathbf{n}\right)\left(\mathbf{e}_i^H\mathbf{n}\right)^H\right] \tag{3.85}$$

$$= \mathbf{e}_i^H\mathbf{C}_{\mathbf{nn}}\mathbf{e}_i. \tag{3.86}$$

The vector $\mathbf{e}_i$ that minimizes this cost function and that produces unbiased estimates is the solution of the constrained optimization problem

$$\mathbf{e}_{\mathrm{B},i} = \arg\min_{\mathbf{e}_i} \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \mathbf{H} = \mathbf{b}_i^H, \tag{3.87}$$

where the index B indicates the BLUE. Solving this constrained optimization problem using the Lagrange multiplier method described in Section 2.4 leads to the BLUE for α_i according to

$$\hat{\alpha}_{\mathrm{B},i} = \mathbf{e}_{\mathrm{B},i}^H \mathbf{y} + c_i \tag{3.88}$$

$$= \mathbf{b}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{y} + c_i. \tag{3.89}$$

Since $\mathbf{b}_i^H$ and c_i are the only terms that depend on the index i, the vector estimator immediately follows as

$$\hat{\boldsymbol{\alpha}}_{\mathrm{B}} = \mathbf{B}\left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{y} + \mathbf{c} \tag{3.90}$$

$$= \mathbf{B}\hat{\mathbf{x}}_{\mathrm{B}} + \mathbf{c}. \tag{3.91}$$

□

Commutation Analysis for the BWLUE

The commutation of the BLUE over linear transformations was analyzed for two cases. In this section, we derive the corresponding results for the BWLUE. In addition, widely linear transformations are considered, which include linear transformations as a special case.

We begin with the first case and assume a linear model of the form as in (3.1) with the difference that the noise is assumed to be improper with known augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$. Then, the BWLUE is given by (3.63). Let the new parameter vector $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\mathbf{x}}}$ be the result of a widely linear transformation of the form

$$\boldsymbol{\alpha} = \mathbf{B}_1 \mathbf{x} + \mathbf{B}_2 \mathbf{x}^* + \mathbf{c}, \tag{3.92}$$

with $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\mathbf{x}}}$ and $\mathbf{c} \in \mathbb{C}^{N_{\mathbf{x}}}$. The augmented notation of (3.92) is

$$\underline{\boldsymbol{\alpha}} = \underline{\mathbf{B}}\,\underline{\mathbf{x}} + \underline{\mathbf{c}}, \tag{3.93}$$

where

$$\underline{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_2^* & \mathbf{B}_1^* \end{bmatrix}, \quad \underline{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \mathbf{c}^* \end{bmatrix}, \tag{3.94}$$

and where $\underline{\mathbf{B}}$ is invertible. The remaining derivation is executed in Appendix A and leads to the result that the BWLUE for $\boldsymbol{\alpha}$ is given in augmented notation as

$$\underline{\hat{\boldsymbol{\alpha}}}_{\mathrm{BW}} = \underline{\mathbf{B}}\,\underline{\hat{\mathbf{x}}}_{\mathrm{BW}} + \underline{\mathbf{c}}, \tag{3.95}$$

where $\underline{\hat{\mathbf{x}}}_{\mathrm{BW}}$ is the BWLUE for $\underline{\mathbf{x}}$ according to (3.63). □

We now turn to the second case and assume $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$, $N_{\boldsymbol{\alpha}} < N_{\mathbf{x}}$ as well as invertability of $\underline{\mathbf{B}}\left(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\right)^{-1} \underline{\mathbf{B}}^H$. The derivation is executed in Appendix B. The result of this derivation proves that the BWLUE for $\boldsymbol{\alpha}$ is given by (3.95), where $\hat{\underline{\mathbf{x}}}_{\mathrm{BW}}$ is the BWLUE for $\underline{\mathbf{x}}$.

3.2 Estimation in a Linear Subspace

Previously, it was shown that the BLUE commutes over linear (actually affine) transformations for two cases. A related situation is investigated in this section, where we assume the parameter vector $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_{\mathbf{x}}}$. We consider the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.96}$$

with the additional model knowledge that $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_{\mathbf{x}}}$ spanned by the columns of a full column rank matrix $\mathbf{B} \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\boldsymbol{\alpha}}}$ such that

$$\mathbf{x} = \mathbf{B}\boldsymbol{\alpha} \tag{3.97}$$

with $N_{\mathbf{x}} > N_{\boldsymbol{\alpha}}$. We seek for the BLUE for $\mathbf{x}$. In this section, $N_{\mathbf{y}} < N_{\mathbf{x}}$ shall be allowed as long as $N_{\mathbf{y}} > N_{\boldsymbol{\alpha}}$ and full column rank of $\mathbf{HB}$ are fulfilled.

3.2.1 Estimation in a Linear Subspace Using the BLUE

With (3.97), the linear model can be rewritten as

$$\mathbf{y} = \mathbf{H}\mathbf{B}\boldsymbol{\alpha} + \mathbf{n}. \tag{3.98}$$

Since we assumed $N_{\mathbf{y}} > N_{\boldsymbol{\alpha}}$ and full column rank of $\mathbf{HB}$, the BLUE for $\boldsymbol{\alpha}$ follows as[3]

$$\hat{\boldsymbol{\alpha}}_{\mathrm{B}} = \left(\mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{y}, \tag{3.99}$$

with

$$\mathbf{C}_{\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}},\mathrm{B}} = \left(\mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\mathbf{B}\right)^{-1}. \tag{3.100}$$

Intuition tells us that $\hat{\mathbf{x}} = \mathbf{B}\hat{\boldsymbol{\alpha}}_{\mathrm{B}}$ is a meaningful estimator, but is it really the BLUE? In the following, we formally show that this estimator is in fact the BLUE.

We seek for the BLUE for $\mathbf{x}$, which uses the additional information in (3.97). Note that the BLUE in (3.48) does not assume any additional constraints on $\mathbf{x}$. Consequently, it is not the true BLUE for $\mathbf{x}$ any more in this situation. To not confuse the reader with varying notations, we refer to (3.48) as the *ordinary* BLUE. For deriving the true

[3] A discussion about the relation between $N_{\mathbf{x}}$ and $N_{\mathbf{y}}$ is presented after the derivation.

BLUE, we need a variance to be minimized and an unbiased constraint. The unbiased constraint can be derived as

$$E_{\mathbf{y}}[\hat{\mathbf{x}}] = \mathbf{x} \tag{3.101}$$
$$E_{\mathbf{y}}[\mathbf{E}\mathbf{y}] = \mathbf{x} \tag{3.102}$$
$$E_{\mathbf{n}}[\mathbf{E}\mathbf{H}\mathbf{x} + \mathbf{E}\mathbf{n}] = \mathbf{x} \tag{3.103}$$
$$E_{\mathbf{n}}[\mathbf{E}\mathbf{H}\mathbf{B}\boldsymbol{\alpha} + \mathbf{E}\mathbf{n}] = \mathbf{B}\boldsymbol{\alpha} \tag{3.104}$$
$$\mathbf{E}\mathbf{H}\mathbf{B}\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\alpha}, \tag{3.105}$$

which directly leads to the unbiased constraint

$$\mathbf{E}\mathbf{H}\mathbf{B} = \mathbf{B}. \tag{3.106}$$

The ordinary BLUE for $\mathbf{x}$ in contrast fulfills the constraint $\mathbf{E}\mathbf{H} = \mathbf{I}$. The additional information allows us to utilize the modified unbiased condition in (3.106). We now analyze and compare these two constraints:

- If $\mathbf{B}$ were invertible, both constraints would be equivalent.
- $\mathbf{E}\mathbf{H} = \mathbf{I}$ has $N_{\mathbf{y}}N_{\mathbf{x}}$ degrees of freedom, which is the number of elements in $\mathbf{E}$. It furthermore has $N_{\mathbf{x}}N_{\mathbf{x}}$ scalar constraints, which is the number of elements in $\mathbf{I}$.
- $\mathbf{E}\mathbf{H}\mathbf{B} = \mathbf{B}$ also has $N_{\mathbf{y}}N_{\mathbf{x}}$ degrees of freedom, but only $N_{\mathbf{x}}N_{\boldsymbol{\alpha}}$ scalar constraints.

Since $N_{\mathbf{x}} > N_{\boldsymbol{\alpha}}$, the modified unbiased constraint in (3.106) is less stringent compared to $\mathbf{E}\mathbf{H} = \mathbf{I}$.

We use the following notation: The i^{th} row of $\mathbf{B}$ is denoted as $\mathbf{b}_i^H$ and the i^{th} row of $\mathbf{E}$ is denoted as $\mathbf{e}_i^H$ such that

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^H \\ \mathbf{b}_2^H \\ \vdots \\ \mathbf{b}_{N_{\mathbf{x}}}^H \end{bmatrix}, \qquad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1^H \\ \mathbf{e}_2^H \\ \vdots \\ \mathbf{e}_{N_{\mathbf{x}}}^H \end{bmatrix}. \tag{3.107}$$

From (3.106), the unbiased constraint for $\mathbf{e}_i^H$ can be extracted and leads to

$$\mathbf{e}_i^H\mathbf{H}\mathbf{B} = \mathbf{b}_i^H. \tag{3.108}$$

The cost function to be minimized is the variance of $\hat{x}_i$, which follows as

$$J(\mathbf{e}_i) = E_{\mathbf{y}}\Big[(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])^H\Big] \tag{3.109}$$
$$= E_{\mathbf{y}}\Big[\big(\mathbf{e}_i^H\mathbf{y} - E_{\mathbf{y}}\big[\mathbf{e}_i^H\mathbf{y}\big]\big)\big(\mathbf{e}_i^H\mathbf{y} - E_{\mathbf{y}}\big[\mathbf{e}_i^H\mathbf{y}\big]\big)^H\Big] \tag{3.110}$$
$$= E_{\mathbf{n}}\Big[\big(\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} - \mathbf{e}_i^H\mathbf{H}\mathbf{x}\big)\big(\mathbf{e}_i^H\mathbf{H}\mathbf{x} + \mathbf{e}_i^H\mathbf{n} - \mathbf{e}_i^H\mathbf{H}\mathbf{x}\big)^H\Big] \tag{3.111}$$
$$= E_{\mathbf{n}}\Big[\big(\mathbf{e}_i^H\mathbf{n}\big)\big(\mathbf{e}_i^H\mathbf{n}\big)^H\Big] \tag{3.112}$$
$$= \mathbf{e}_i^H\mathbf{C}_{\mathbf{n}\mathbf{n}}\mathbf{e}_i. \tag{3.113}$$

In summary, the following constrained optimization problem is obtained:

$$\mathbf{e}_{\mathrm{B},i} = \arg \min_{\mathbf{e}_i} \ \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \mathbf{H} \mathbf{B} = \mathbf{b}_i^H. \tag{3.114}$$

The Lagrangian cost function for this problem is given by

$$\mathcal{L}(\mathbf{e}_i) = \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{e}_i + \boldsymbol{\lambda}^H \left(\mathbf{B}^H \mathbf{H}^H \mathbf{e}_i - \mathbf{b}_i\right) + \boldsymbol{\lambda}^T \left(\mathbf{B}^T \mathbf{H}^T \mathbf{e}_i^* - \mathbf{b}_i^*\right). \tag{3.115}$$

The Wirtinger derivative with respect to $\mathbf{e}_i$ produces

$$\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} = \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} + \boldsymbol{\lambda}^H \mathbf{B}^H \mathbf{H}^H. \tag{3.116}$$

Setting (3.116) equal to zero results in

$$\mathbf{e}_{\mathrm{B},i}^H = -\boldsymbol{\lambda}^H \mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}. \tag{3.117}$$

Inserting (3.117) into the constraint in (3.108) produces

$$-\boldsymbol{\lambda}^H \mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H} \mathbf{B} = \mathbf{b}_i^H \tag{3.118}$$

$$-\boldsymbol{\lambda}^H = \mathbf{b}_i^H \left(\mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H} \mathbf{B}\right)^{-1}. \tag{3.119}$$

Reinserting this result into the expression for $\mathbf{e}_i^H$ in (3.117) yields

$$\mathbf{e}_{\mathrm{B},i}^H = \mathbf{b}_i^H \left(\mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}. \tag{3.120}$$

Since $\mathbf{b}_i^H$ is the only term in (3.120) that depends on the index i, the expression for the estimator matrix immediately follows as

$$\mathbf{E}_{\mathrm{B}} = \begin{bmatrix} \mathbf{e}_{\mathrm{B},1}^H \\ \mathbf{e}_{\mathrm{B},2}^H \\ \vdots \\ \mathbf{e}_{\mathrm{B},N_{\mathbf{x}}}^H \end{bmatrix} = \mathbf{B} \left(\mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}. \tag{3.121}$$

Comparing this result with (3.99) proofs that

$$\hat{\mathbf{x}}_{\mathrm{B}} = \mathbf{B} \hat{\boldsymbol{\alpha}}_{\mathrm{B}} \tag{3.122}$$

holds. From (3.122) it follows that the covariance matrix of $\hat{\mathbf{x}}_{\mathrm{B}}$ is given by

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{B}} = \mathbf{B} \mathbf{C}_{\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}},\mathrm{B}} \mathbf{B}^H. \tag{3.123}$$

Recall that we assumed that $\mathbf{HB}$ has full column rank, $N_{\mathbf{y}} > N_{\boldsymbol{\alpha}}$ and $N_{\mathbf{x}} > N_{\boldsymbol{\alpha}}$, but no additional assumption on the relation between $N_{\mathbf{x}}$ and $N_{\mathbf{y}}$ has been made. Hence, this result is applicable even for $N_{\mathbf{x}} > N_{\mathbf{y}}$ as long as the mentioned assumptions hold.

The results of these investigations are summarized in

Result 3.1 (Estimation in a Linear Subspace Using the BLUE)

Consider the linear model in (3.1), where $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ is the measurement vector, $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a known measurement matrix, and $\mathbf{n} \in \mathbb{C}^{N_\mathbf{y}}$ is a zero mean random noise vector with known covariance matrix $\mathbf{C}_{\mathbf{nn}}$. If it is known that $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_\mathbf{x}}$ spanned by the columns of a full column rank matrix $\mathbf{B} \in \mathbb{C}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}}$ with $N_\mathbf{x} > N_{\boldsymbol{\alpha}}$ according to (3.97), and if

- $N_\mathbf{y} > N_{\boldsymbol{\alpha}}$, and
- $\mathbf{HB}$ has full column rank,

then the BLUE for $\mathbf{x}$ is given by (3.122) where $\hat{\boldsymbol{\alpha}}_\mathrm{B}$ is the BLUE for $\boldsymbol{\alpha}$ according to (3.99). This estimator is unbiased in the classical sense, i.e., it fulfills $E_\mathbf{y}[\hat{\mathbf{x}}_\mathrm{B}] = \mathbf{x}$, and its covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{B}}$ is given by (3.123).

In the following, two possible applications and their connection are discussed.

Application 1: System Identification via Frequency Domain Measurements

Let the parameter vector $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ be samples of the frequency response (i.e. Fourier coefficients) of a linear time-invariant (LTI) system. Measurements of this frequency response are possible for some but not necessarily all frequencies. E.g., DC measurements are sometimes difficult to conduct as it is the case for ultrasonic measurements. The vector of possible measurements is denoted as $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ with $N_\mathbf{y} < N_\mathbf{x}$. Consequently, $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a selection matrix. Let $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$ denote the impulse response of the system with $N_\mathbf{y} > N_{\boldsymbol{\alpha}}$ such that $\mathbf{B} \in \mathbb{C}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}}$ is given by the first $N_{\boldsymbol{\alpha}}$ columns of a discrete Fourier transform (DFT) matrix of size $N_\mathbf{x} \times N_\mathbf{x}$. Finally, we end up with the model

$$\begin{aligned} \mathbf{y} &= \mathbf{Hx} + \mathbf{n} && (3.124) \\ &= \mathbf{HB}\boldsymbol{\alpha} + \mathbf{n}. && (3.125) \end{aligned}$$

If $\mathbf{HB}$ has full column rank Result 3.1 is applicable and the full frequency response can be estimated.

Application 2: System Identification via Time Domain Measurements

Again, $\boldsymbol{\alpha}$ denotes the impulse response of an unknown LTI system. $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ represents time domain measurements of the output of the system for a given input signal, modelled as

$$\mathbf{y} = \mathbf{H}'\boldsymbol{\alpha} + \mathbf{n}. \quad (3.126)$$

Here, $\mathbf{H}' \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a convolution matrix generated with the samples of the input signal. Furthermore, let $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ be the frequency response of the LTI system generated by the DFT of the zero padded impulse response. This can be describes via

$$\mathbf{x} = \mathbf{B}\boldsymbol{\alpha}, \tag{3.127}$$

where $\mathbf{B} \in \mathbb{C}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}}$ is given by the first $N_{\boldsymbol{\alpha}}$ columns of a DFT matrix of size $N_\mathbf{x} \times N_\mathbf{x}$ with $N_\mathbf{x} > N_{\boldsymbol{\alpha}}$. This problem looks more like the ones discussed in Section 3.1.2 with the major difference that $\mathbf{B}$ has more rows than columns.

We denote the inverse DFT matrix of size $N_\mathbf{x} \times N_\mathbf{x}$ as $\mathbf{F}^{-1}$, and introduce

$$\mathbf{W} = \begin{bmatrix} \mathbf{I}^{N_{\boldsymbol{\alpha}} \times N_{\boldsymbol{\alpha}}} & \mathbf{0}^{N_{\boldsymbol{\alpha}} \times N_\mathbf{x} - N_{\boldsymbol{\alpha}}} \end{bmatrix}, \tag{3.128}$$

then

$$\boldsymbol{\alpha} = \mathbf{W}\mathbf{F}^{-1}\mathbf{x} \tag{3.129}$$

$$= \underbrace{\mathbf{W}\mathbf{F}^{-1}\mathbf{B}}_{\mathbf{I}}\boldsymbol{\alpha}. \tag{3.130}$$

Combining this expression with the model in (3.126) allows for

$$\mathbf{y} = \underbrace{\mathbf{H}'\mathbf{W}\mathbf{F}^{-1}}_{\mathbf{H}}\mathbf{B}\boldsymbol{\alpha} + \mathbf{n} \tag{3.131}$$

$$= \mathbf{H}\underbrace{\mathbf{B}\boldsymbol{\alpha}}_{\mathbf{x}} + \mathbf{n} \tag{3.132}$$

$$= \mathbf{H}\mathbf{x} + \mathbf{n}. \tag{3.133}$$

Since $\mathbf{HB}$ has full column rank and $N_\mathbf{y} > N_{\boldsymbol{\alpha}}$, Result 3.1 is applicable and the full frequency response can be estimated.

3.2.2 Estimation in a Linear Subspace Using the BWLUE

An extension of Section 3.2.1 to the BWLUE is presented in this section. We again consider the linear model in (3.1) with the difference that $\mathbf{n}$ may now be a zero mean improper noise vector. The BWLUE to be derived in the following incorporates the additional knowledge that $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_\mathbf{x}}$ according to

$$\mathbf{x} = \mathbf{B}\boldsymbol{\alpha}, \tag{3.134}$$

where $\mathbf{B} \in \mathbb{C}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}}$ and $N_\mathbf{x} > N_{\boldsymbol{\alpha}}$. We again assume $N_\mathbf{y} > N_{\boldsymbol{\alpha}}$ and full column rank of $\mathbf{HB}$ such that the BWLUE for $\boldsymbol{\alpha}$ follows as

$$\underline{\hat{\boldsymbol{\alpha}}}_{\mathrm{BW}} = \left(\underline{\mathbf{B}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\,\underline{\mathbf{B}}\right)^{-1}\underline{\mathbf{B}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}}, \tag{3.135}$$

where

$$\underline{\mathbf{B}} = \begin{bmatrix} \mathbf{B} & \mathbf{0}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}} \\ \mathbf{0}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}} & \mathbf{B}^* \end{bmatrix}. \tag{3.136}$$

We seek for the BWLUE for $\mathbf{x}$, which uses the additional model knowledge in (3.134). Clearly, the BWLUE has to fulfill $E_{\mathbf{y}}[\hat{\mathbf{x}}] = \mathbf{x}$, which in augmented notation reads as

$$
\begin{aligned}
E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}] &= \underline{\mathbf{x}} && (3.137)\\
E_{\mathbf{y}}[\underline{\mathbf{E}}\,\underline{\mathbf{y}}] &= \underline{\mathbf{x}} && (3.138)\\
E_{\mathbf{n}}[\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{E}}\,\underline{\mathbf{n}}] &= \underline{\mathbf{x}} && (3.139)\\
E_{\mathbf{n}}[\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{B}}\,\underline{\boldsymbol{\alpha}} + \underline{\mathbf{E}}\,\underline{\mathbf{n}}] &= \underline{\mathbf{B}}\,\underline{\boldsymbol{\alpha}} && (3.140)\\
\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{B}}\,\underline{\boldsymbol{\alpha}} &= \underline{\mathbf{B}}\,\underline{\boldsymbol{\alpha}}, && (3.141)
\end{aligned}
$$

from which the unbiased constraint follows as

$$\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{B}} = \underline{\mathbf{B}}. \tag{3.142}$$

The ordinary BWLUE for $\mathbf{x}$ fulfills the constraint $\underline{\mathbf{E}}\,\underline{\mathbf{H}} = \mathbf{I}$. Comparing this with the modified constraint in (3.142) allows similar statements as in Section 3.2.1

- If $\underline{\mathbf{B}}$ would be invertible, both constraints would be equivalent.

- $\underline{\mathbf{E}}\,\underline{\mathbf{H}} = \mathbf{I}$ has $2N_{\mathbf{y}}N_{\mathbf{x}}$ degrees of freedom, which is the number of elements in $\mathbf{F}$ and $\mathbf{G}$ within $\underline{\mathbf{E}}$ according to (2.23). Furthermore, it has $2N_{\mathbf{x}}N_{\mathbf{x}}$ scalar constraints, which is the number of elements in the upper half of $\mathbf{I}$.

- $\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{B}} = \underline{\mathbf{B}}$ also has $2N_{\mathbf{y}}N_{\mathbf{x}}$ degrees of freedom but only $2N_{\mathbf{x}}N_{\boldsymbol{\alpha}}$ scalar constraints, which corresponds to the number of elements in the upper half of $\underline{\mathbf{B}}$.

Since $N_{\mathbf{x}} > N_{\boldsymbol{\alpha}}$, the modified unbiased constraint in (3.142) is less stringent compared to $\underline{\mathbf{E}}\,\underline{\mathbf{H}} = \mathbf{I}$.

For the derivation, we utilize a similar notation as in Section 3.2.1. Let the i^{th} row of $\underline{\mathbf{B}}$ be denoted as $\mathbf{b}_i^H$ and let the i^{th} row of $\underline{\mathbf{E}}$ be denoted as $\mathbf{e}_i^H$

$$\underline{\mathbf{B}} = \begin{bmatrix} \mathbf{b}_1^H \\ \mathbf{b}_2^H \\ \vdots \\ \mathbf{b}_{N_{\mathbf{x}}}^H \end{bmatrix}, \qquad \underline{\mathbf{E}} = \begin{bmatrix} \mathbf{e}_1^H \\ \mathbf{e}_2^H \\ \vdots \\ \mathbf{e}_{N_{\boldsymbol{\alpha}}}^H \end{bmatrix}. \tag{3.143}$$

From (3.142), the unbiased constraint for $\mathbf{e}_i^H$ follows as

$$\mathbf{e}_i^H\underline{\mathbf{H}}\,\underline{\mathbf{B}} = \mathbf{b}_i^H. \tag{3.144}$$

As cost function the variance of $\hat{x}_i$ is used, which reads as

$$
\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y}}\left[(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])^H\right] && (3.145)\\
&= E_{\mathbf{y}}\left[\left(\mathbf{e}_i^H\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\mathbf{e}_i^H\underline{\mathbf{y}}\right]\right)\left(\mathbf{e}_i^H\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\mathbf{e}_i^H\underline{\mathbf{y}}\right]\right)^H\right] && (3.146)\\
&= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H\underline{\mathbf{n}} - \mathbf{e}_i^H\underline{\mathbf{H}}\,\underline{\mathbf{x}}\right)\left(\mathbf{e}_i^H\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H\underline{\mathbf{n}} - \mathbf{e}_i^H\underline{\mathbf{H}}\,\underline{\mathbf{x}}\right)^H\right] && (3.147)\\
&= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H\underline{\mathbf{n}}\right)\left(\mathbf{e}_i^H\underline{\mathbf{n}}\right)^H\right] && (3.148)\\
&= \mathbf{e}_i^H\underline{\mathbf{C}}_{\mathbf{nn}}\mathbf{e}_i. && (3.149)
\end{aligned}
$$

By combining (3.144) and (3.149) the constrained optimization problem is given by

$$\mathbf{e}_{\mathrm{BW},i} = \arg\min_{\mathbf{e}_i} \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{B}} = \mathbf{b}_i^H. \tag{3.150}$$

Therewith, the Lagrangian cost function for this optimization problem yields

$$\mathcal{L}(\mathbf{e}_i) = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i + \boldsymbol{\lambda}^H \big(\underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \mathbf{e}_i - \mathbf{b}_i\big) + \boldsymbol{\lambda}^T \big(\underline{\mathbf{B}}^T \underline{\mathbf{H}}^T \mathbf{e}_i^* - \mathbf{b}_i^*\big). \tag{3.151}$$

Setting the Wirtinger derivative of (3.151) w.r.t. $\mathbf{e}_i$ equal to zero allows identifying $\mathbf{e}_{\mathrm{BW},i}$ as

$$\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} + \boldsymbol{\lambda}^H \underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \overset{!}{=} \mathbf{0} \tag{3.152}$$

$$\mathbf{e}_{\mathrm{BW},i}^H = -\boldsymbol{\lambda}^H \underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.153}$$

Inserting (3.153) into the constraint in (3.144) results in

$$-\boldsymbol{\lambda}^H \underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{B}} = \mathbf{b}_i^H, \tag{3.154}$$

and after rearranging we have

$$-\boldsymbol{\lambda}^H = \mathbf{b}_i^H \big(\underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{B}}\big)^{-1}. \tag{3.155}$$

Reinserting this result into the expression for $\mathbf{e}_i^H$ in (3.153) yields

$$\mathbf{e}_{\mathrm{BW},i}^H = \mathbf{b}_i^H \big(\underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{B}}\big)^{-1} \underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.156}$$

Since $\mathbf{b}_i^H$ is the only term in (3.156) that depends on the index i, the expression for the estimator matrix $\underline{\mathbf{E}}_{\mathrm{BW}}$ immediately follows as

$$\underline{\mathbf{E}}_{\mathrm{BW}} = \underline{\mathbf{B}} \big(\underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{B}}\big)^{-1} \underline{\mathbf{B}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.157}$$

Comparing this result to (3.135) proofs that

$$\hat{\underline{\mathbf{x}}}_{\mathrm{BW}} = \underline{\mathbf{B}}\,\hat{\underline{\boldsymbol{\alpha}}}_{\mathrm{BW}} \tag{3.158}$$

holds. Further, from (3.158), it follows that the augmented covariance matrix of $\hat{\mathbf{x}}_{\mathrm{BW}}$ is given by

$$\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{BW}} = \underline{\mathbf{B}}\,\underline{\mathbf{C}}_{\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}},\mathrm{BW}} \underline{\mathbf{B}}^H. \tag{3.159}$$

Note that the only assumption necessary for deriving this result was that the matrix product $\mathbf{H}\,\mathbf{B}$ has full column rank and $N_{\mathbf{y}} > N_{\boldsymbol{\alpha}}$. This also allows parameter vectors with $N_{\mathbf{x}} > N_{\mathbf{y}}$ as long as the mentioned assumptions hold.

The derived results are summarized in

Result 3.2 (Estimation in a Linear Subspace Using the BWLUE)

Consider the linear model in (3.1) where $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ is the measurement vector, $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a known measurement matrix, and $\mathbf{n} \in \mathbb{C}^{N_\mathbf{y}}$ is a zero mean random proper or improper noise vector with known augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$. If it is known that $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_\mathbf{x}}$ with $N_\mathbf{x} > N_{\boldsymbol{\alpha}}$ according to (3.134), and if

- $N_\mathbf{y} > N_{\boldsymbol{\alpha}}$, and
- $\mathbf{H}\,\mathbf{B}$ has full column rank,

then the BWLUE for $\mathbf{x}$ is given in augmented notation by (3.158), where $\underline{\hat{\boldsymbol{\alpha}}}_{\mathrm{BW}}$ is the BWLUE for $\boldsymbol{\alpha}$ according to (3.135). This estimator is unbiased in the classical sense, i.e., it fulfills $E_\mathbf{y}[\underline{\hat{\mathbf{x}}}_{\mathrm{BW}}] = \underline{\mathbf{x}}$, and its augmented covariance matrix $\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{BW}}$ is given by (3.159).

We now shortly discuss an extension of Result 3.2. When replacing the linear equation in (3.134) with the widely linear transformation

$$\mathbf{x} = \mathbf{B}_1\boldsymbol{\alpha} + \mathbf{B}_2\boldsymbol{\alpha}^*, \tag{3.160}$$

the same result would have been obtained as long as the product $\underline{\mathbf{H}}\,\underline{\mathbf{B}}$ has full column rank, where

$$\underline{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_2^* & \mathbf{B}_1^* \end{bmatrix}. \tag{3.161}$$

3.2.3 Estimation in a Linear Subspace Using the LS Estimator

Besides the linear model assumption, we assume to have the additional knowledge that $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ lies in a linear subspace of $\mathbb{C}^{N_\mathbf{x}}$ spanned by the columns of a full column rank matrix $\mathbf{B} \in \mathbb{C}^{N_\mathbf{x} \times N_{\boldsymbol{\alpha}}}$ such that

$$\mathbf{x} = \mathbf{B}\boldsymbol{\alpha} \tag{3.162}$$

with $N_\mathbf{x} > N_{\boldsymbol{\alpha}}$. We already derived the BLUE and the BWLUE for this task. For the LS estimator, however, the incorporation of (3.162) into the derivation of the LS estimator is generally not straightforward.

To overcome this, we transform (3.162) into an equivalent constraint of the form $\mathbf{A}\mathbf{x} = \mathbf{0}$, which corresponds to the constraint that $\mathbf{x}$ lies in the nullspace of $\mathbf{A}$. In fact, it is possible to find a matrix $\mathbf{A}$ whose nullspace equals the space spanned by the columns of $\mathbf{B}$. This is shown in the following. First, the nullspace of $\mathbf{B}^H$ is identified. Let this nullspace have the dimension $N_0 = N_\mathbf{x} - N_{\boldsymbol{\alpha}}$ and let $\mathbf{N} \in \mathbb{C}^{N_\mathbf{x} \times N_0}$ be a matrix whose columns span the nullspace of $\mathbf{B}^H$. Consequently,

$$\mathbf{B}^H\mathbf{N} = \mathbf{0}^{N_{\boldsymbol{\alpha}}\times N_0} \tag{3.163}$$

holds. Taking the conjugate complex transpose of (3.163) and defining the matrix $\mathbf{A} = \mathbf{N}^H \in \mathbb{C}^{N_0\times N_{\mathbf{x}}}$ produces

$$\mathbf{AB} = \mathbf{0}^{N_0\times N_{\boldsymbol{\alpha}}}, \tag{3.164}$$

Consequently, the columns of $\mathbf{B}$ span the nullspace of $\mathbf{A}$. Furthermore, the set of solutions of

$$\mathbf{Ax} = \mathbf{0}^{N_0\times 1} \tag{3.165}$$

corresponds to (3.162) with arbitrary $\boldsymbol{\alpha}$. Hence, the information about $\mathbf{x}$ lying in a subspace of $\mathbb{C}^{N_{\mathbf{x}}}$ spanned by the columns of $\mathbf{B}$ has been transformed into the constraint in (3.165). This allows utilizing the constrained LS estimator derived in Section 3.1. This constrained LS estimator produces estimates that lie in a subspace of $\mathbb{C}^{N_{\mathbf{x}}}$ spanned by the columns of $\mathbf{B}$. Hence, we obtain the following result:

Result 3.3 (Estimation in a Linear Subspace Using the LS Estimator)

Consider the linear model in (3.1), where $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ is the measurement vector, $\mathbf{H} \in \mathbb{C}^{N_{\mathbf{y}}\times N_{\mathbf{x}}}$ is a known measurement matrix, and $\mathbf{n} \in \mathbb{C}^{N_{\mathbf{y}}}$ is a zero mean random noise vector. It shall be known that $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_{\mathbf{x}}}$ spanned by the columns of a full column rank matrix $\mathbf{B} \in \mathbb{C}^{N_{\mathbf{x}}\times N_{\boldsymbol{\alpha}}}$ with $N_{\mathbf{x}} > N_{\boldsymbol{\alpha}}$ according to (3.97). Let the nullspace of $\mathbf{B}^H$ have the dimension N_0 and let $\mathbf{N} \in \mathbb{C}^{N_{\mathbf{x}}\times N_0}$ be the matrix whose columns span the nullspace of $\mathbf{B}^H$. Then, the LS estimator that incorporates the knowledge that $\mathbf{x}$ lies in a linear subspace of $\mathbb{C}^{N_{\mathbf{x}}}$ spanned by the columns of matrix $\mathbf{B}$ is given by the constrained LS estimator in (3.29) with $\mathbf{A} = \mathbf{N}^H$ and $\mathbf{b} = \mathbf{0}^{N_0\times 1}$. This estimator is unbiased in the classical sense, i.e., it fulfills $E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}] = \mathbf{x}$, and its covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}}$ is given by (3.35).

3.3 Estimation with Additional Constraints on the Parameter Vector

In many practical examples, more model knowledge than the measurement matrix $\mathbf{H}$ and the noise statistics $\mathbf{C}_{\mathbf{nn}}$ is available. As an example, some physical systems are known to be unable to transmit any DC signals. Such a physical system, e.g., could be a communication channel or a sensor with a differential measurement principle. If the system can be described by an impulse response, then the inability to transmit DC signals corresponds to the integral of the impulse response being zero. We assume the samples of the discrete-time impulse response $x[n]$ sum up to zero such that

$$\sum_{n=0}^{N_{\mathbf{x}}-1} x[n] = 0, \tag{3.166}$$

and we seek for an estimator that produces estimates that fulfill

$$\sum_{n=0}^{N_{\mathbf{x}}-1} \hat{x}[n] = 0. \tag{3.167}$$

The vector notation of the constraint in (3.166) corresponds to $\mathbf{1}^T\mathbf{x} = 0$, where $\mathbf{1}$ is a column vector with length $N_{\mathbf{x}}$ with all elements being 1. We introduce the even more general constraint

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{3.168}$$

with full row rank $\mathbf{A} \in \mathbb{C}^{N_{\mathbf{b}} \times N_{\mathbf{x}}}$, $\mathbf{b} \in \mathbb{C}^{N_{\mathbf{b}}}$, $N_{\mathbf{b}} < N_{\mathbf{x}}$, and where $\mathbf{x} \in \mathbb{C}^{N_{\mathbf{x}}}$ can be a general parameter vector rather than only an impulse response. Eq. (3.168) corresponds to (3.166) for $\mathbf{A} = \mathbf{1}^T$ and $\mathbf{b} = 0$ but it also allows incorporating other types of model knowledge.

We seek for estimators that incorporate knowledge about the constraints in (3.168). In other words, we seek estimators that fulfill

$$\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{3.169}$$

A possible estimator for this task is the constrained LS estimator discussed in Section 3.1. However, to the best of our knowledge constrained versions of the BLUE and the BWLUE have not been published so far. In the following two sections, these novel estimators are proposed. It will turn out that these estimators allow for $N_{\mathbf{y}} < N_{\mathbf{x}}$, which is forbidden for the constrained LS estimator. A detailed discussion about this will be presented.

3.3.1 Constrained BLUE

We assume the linear model in (3.1) holds. In coherence with the constrained LS estimator in (3.29), we assume the estimator to be affine and of the form

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{f}. \tag{3.170}$$

As the estimator is actually affine the term 'linear' in the abbreviation 'BLUE' might be somewhat misleading. However, since also for other affine estimators the term 'linear' is usually used, we call the estimator constrained BLUE. The goal is now to find the estimator matrix $\mathbf{E} \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\mathbf{y}}}$ and the vector $\mathbf{f} \in \mathbb{C}^{N_{\mathbf{x}}}$. The constrained BLUE has to fulfill two types of constraints. The first one is the unbiased constraint

$$\begin{aligned} E_{\mathbf{y}}[\hat{\mathbf{x}}] &= E_{\mathbf{y}}[\mathbf{E}\mathbf{y} + \mathbf{f}] && (3.171)\\ &= E_{\mathbf{n}}[\mathbf{E}(\mathbf{H}\mathbf{x} + \mathbf{n}) + \mathbf{f}] && (3.172)\\ &= \mathbf{E}\mathbf{H}\mathbf{x} + \mathbf{f} \stackrel{!}{=} \mathbf{x}. && (3.173) \end{aligned}$$

By letting $\mathbf{e}_i^H$ be the i^{th} row of $\mathbf{E}$, x_i be the i^{th} element of $\mathbf{x}$, and f_i be the i^{th} element of $\mathbf{f}$ such that

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1^H \\ \vdots \\ \mathbf{e}_{N_\mathbf{x}}^H \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{N_\mathbf{x}} \end{bmatrix}, \qquad \mathbf{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_{N_\mathbf{x}} \end{bmatrix}, \tag{3.174}$$

the unbiased constraint for $\mathbf{e}_i^H$ can be extracted from (3.173) and is of the form

$$\mathbf{e}_i^H \mathbf{H}\mathbf{x} + f_i \stackrel{!}{=} x_i. \tag{3.175}$$

The second type of constraints are given by (3.169). For each $i \in \{1, 2, ..., N_\mathbf{x}\}$ the variance of $\hat{x}_i$ serves as a cost function which is a function of $\mathbf{e}_i$ given as

$$\begin{aligned} J(\mathbf{e}_i) &= E_\mathbf{y}\left[(\hat{x}_i - E_\mathbf{y}[\hat{x}_i])(\hat{x}_i - E_\mathbf{y}[\hat{x}_i])^H\right] && (3.176) \\ &= E_\mathbf{y}\left[\left(\mathbf{e}_i^H\mathbf{y} + f_i - \mathbf{e}_i^H E_\mathbf{y}[\mathbf{y}] - f_i\right)\left(\mathbf{e}_i^H\mathbf{y} + f_i - \mathbf{e}_i^H E_\mathbf{y}[\mathbf{y}] - f_i\right)^H\right] && (3.177) \\ &= E_\mathbf{n}\left[\left(\mathbf{e}_i^H(\mathbf{H}\mathbf{x} + \mathbf{n}) - \mathbf{e}_i^H\mathbf{H}\mathbf{x}\right)\left(\mathbf{e}_i^H(\mathbf{H}\mathbf{x} + \mathbf{n}) - \mathbf{e}_i^H\mathbf{H}\mathbf{x}\right)^H\right] && (3.178) \\ &= E_\mathbf{n}\left[\left(\mathbf{e}_i^H\mathbf{n}\right)\left(\mathbf{e}_i^H\mathbf{n}\right)^H\right] && (3.179) \\ &= \mathbf{e}_i^H \mathbf{C}_{\mathbf{n}\mathbf{n}} \mathbf{e}_i. && (3.180) \end{aligned}$$

We note, that (3.169) represents constraints in $\hat{\mathbf{x}}$, however, the i^{th} cost function is a function of the vector $\mathbf{e}_i$, which is conflicting. We therefore transform the constraints in (3.169) into a different but equivalent form, combine them with (3.173), and finally end up with constraints on $\mathbf{e}_i$.

We start with an analysis of $\mathbf{A}\mathbf{x} = \mathbf{b}$ in (3.168). This linear system of equations has an infinite number of solutions that can be described as

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \ldots + \mathbf{x}_{N_0}\alpha_{N_0}, \tag{3.181}$$

where the vectors $\mathbf{x}_i$, $i = 1, \ldots, N_0$, span the nullspace of $\mathbf{A}$ such that $\mathbf{A}\mathbf{x}_i = \mathbf{0}^{N_\mathbf{b}\times 1}$, N_0 is the dimension of the nullspace of $\mathbf{A}$ with $N_0 = N_\mathbf{x} - N_\mathbf{b}$, the scalar coefficients α_i, $i = 1, \ldots, N_0$ are in general complex-valued and arbitrary, and $\mathbf{x}_p$ is an arbitrary particular solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$, e.g., the minimum norm solution $\mathbf{x}_p = \mathbf{A}^H\left(\mathbf{A}\mathbf{A}^H\right)^{-1}\mathbf{b}$. However, the particular choice of $\mathbf{x}_p$ is not of importance in the following. Eq. (3.181) can be brought into the form

$$\mathbf{x} = \mathbf{x}_p + \mathbf{N}\boldsymbol{\alpha}, \tag{3.182}$$

where

$$\mathbf{N} = \begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_{N_0} \end{bmatrix} \in \mathbb{C}^{N_\mathbf{x}\times N_0}, \qquad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{N_0} \end{bmatrix} \in \mathbb{C}^{N_0}. \tag{3.183}$$

With this notation we have $\mathbf{AN} = \mathbf{0}^{N_\mathbf{b} \times N_0}$. Inserting (3.182) into (3.173) results in

$$E_\mathbf{y}[\hat{\mathbf{x}}] = \mathbf{EH}(\mathbf{x}_p + \mathbf{N}\boldsymbol{\alpha}) + \mathbf{f} \stackrel{!}{=} \mathbf{x}_p + \mathbf{N}\boldsymbol{\alpha} \tag{3.184}$$

$$\Leftrightarrow (\mathbf{EHN} - \mathbf{N})\boldsymbol{\alpha} + (\mathbf{EH} - \mathbf{I})\mathbf{x}_p + \mathbf{f} \stackrel{!}{=} \mathbf{0}. \tag{3.185}$$

To fulfill this equation for every possible vector $\boldsymbol{\alpha}$, we deduce the following two constraints for $\mathbf{E}$ and $\mathbf{f}$:

$$\mathbf{EHN} = \mathbf{N} \tag{3.186}$$

$$\mathbf{f} = (\mathbf{I} - \mathbf{EH})\mathbf{x}_p. \tag{3.187}$$

Let the i^{th} row of $\mathbf{N}$ be denoted as $\mathbf{n}_i^H$, then the constraint for $\mathbf{e}_i^H$ can be extracted from (3.186) and leads to

$$\mathbf{e}_i^H \mathbf{HN} = \mathbf{n}_i^H. \tag{3.188}$$

We are now finally able to formulate the constrained optimization problem for $\mathbf{e}_i$:

$$\mathbf{e}_{\text{CB},i} = \arg \min_{\mathbf{e}_i} \; \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \mathbf{HN} = \mathbf{n}_i^H. \tag{3.189}$$

We solve this constrained optimization problem using the Lagrangian multiplier method. The Lagrangian cost function for this problem is given by

$$\mathcal{L}(\mathbf{e}_i) = \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{e}_i + \boldsymbol{\lambda}^H \left(\mathbf{N}^H \mathbf{H}^H \mathbf{e}_i - \mathbf{n}_i\right) + \boldsymbol{\lambda}^T \left(\mathbf{N}^T \mathbf{H}^T \mathbf{e}_i^* - \mathbf{n}_i^*\right). \tag{3.190}$$

The Wirtinger derivative with respect to $\mathbf{e}_i$ produces

$$\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} = \mathbf{e}_i^H \mathbf{C}_{\mathbf{nn}} + \boldsymbol{\lambda}^H \mathbf{N}^H \mathbf{H}^H. \tag{3.191}$$

Setting (3.191) equal to zero results in

$$\mathbf{e}_{\text{CB},i}^H = -\boldsymbol{\lambda}^H \mathbf{N}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}. \tag{3.192}$$

Assuming full column rank of $\mathbf{HN}$, which implies $N_\mathbf{y} \geq N_0$, and inserting (3.192) into the constraint in (3.189) produces

$$-\boldsymbol{\lambda}^H = \mathbf{n}_i^H \left(\mathbf{N}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{HN}\right)^{-1}. \tag{3.193}$$

Reinserting this result into the expression for $\mathbf{e}_i^H$ in (3.192) yields

$$\mathbf{e}_{\text{CB},i}^H = \mathbf{n}_i^H \left(\mathbf{N}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{HN}\right)^{-1} \mathbf{N}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}. \tag{3.194}$$

Since $\mathbf{n}_i^H$ is the only term in (3.194) that depends on the index i, the expression for the estimator matrix is given by

$$\mathbf{E}_{\text{CB}} = \mathbf{N}\left(\mathbf{N}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{HN}\right)^{-1} \mathbf{N}^H \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}. \tag{3.195}$$

In the following, we denote

$$\mathbf{P} = \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}. \tag{3.196}$$

Inserting (3.187) and (3.195) into (3.170) finally leads to the constrained BLUE in the form of

$$\hat{\mathbf{x}}_{\mathrm{CB}} = \mathbf{E}_{\mathrm{CB}}\mathbf{y} + (\mathbf{I} - \mathbf{E}_{\mathrm{CB}}\mathbf{H})\mathbf{x}_p \tag{3.197}$$

$$= \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} + \Big(\mathbf{I} - \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{P}\Big)\mathbf{x}_p \tag{3.198}$$

$$= \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_p) + \mathbf{x}_p. \tag{3.199}$$

This estimator is unbiased, which can be shown by incorporating (3.182) and (3.186) into (3.197)

$$E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{CB}}] = \mathbf{E}_{\mathrm{CB}}E_{\mathbf{y}}[\underline{\mathbf{y}}] + (\mathbf{I} - \mathbf{E}_{\mathrm{CB}}\mathbf{H})\mathbf{x}_p \tag{3.200}$$

$$= \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{x} + (\mathbf{I} - \mathbf{E}_{\mathrm{CB}}\mathbf{H})(\mathbf{x} - \mathbf{N}\boldsymbol{\alpha}) \tag{3.201}$$

$$= \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{x} + \mathbf{x} - \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{x} - \mathbf{N}\boldsymbol{\alpha} + \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{N}\boldsymbol{\alpha} \tag{3.202}$$

$$= \mathbf{x} - \mathbf{N}\boldsymbol{\alpha} + \underbrace{\mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{N}}_{\mathbf{N}}\boldsymbol{\alpha} \tag{3.203}$$

$$= \mathbf{x}. \tag{3.204}$$

Following similar arguments, it holds that

$$\hat{\mathbf{x}}_{\mathrm{CW}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{CB}}] = \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{x} + \mathbf{E}_{\mathrm{CB}}\mathbf{n} + \mathbf{x} - \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{x} - \mathbf{N}\boldsymbol{\alpha} + \mathbf{E}_{\mathrm{CB}}\mathbf{H}\mathbf{N}\boldsymbol{\alpha} - \mathbf{x} \tag{3.205}$$

$$= \mathbf{E}_{\mathrm{CB}}\mathbf{n}. \tag{3.206}$$

With that, the covariance matrix of $\hat{\mathbf{x}}_{\mathrm{CB}}$ can be derived as

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{CB}} = E_{\mathbf{y}}\Big[(\hat{\mathbf{x}}_{\mathrm{CB}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{CB}}])(\hat{\mathbf{x}}_{\mathrm{CB}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{CB}}])^H\Big] \tag{3.207}$$

$$= \mathbf{E}_{\mathrm{CB}}E_{\mathbf{n}}\big[\mathbf{n}\mathbf{n}^H\big]\mathbf{E}_{\mathrm{CB}}^H \tag{3.208}$$

$$= \mathbf{E}_{\mathrm{CB}}\mathbf{C}_{\mathbf{nn}}\mathbf{E}_{\mathrm{CB}}^H \tag{3.209}$$

$$= \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H \tag{3.210}$$

$$= \mathbf{N}\big(\mathbf{N}^H\mathbf{H}^H\mathbf{P}\mathbf{H}\mathbf{N}\big)^{-1}\mathbf{N}^H, \tag{3.211}$$

where $\hat{\mathbf{x}}_{\mathrm{CB}}$ in (3.199) is actually independent of the particular choice of $\mathbf{x}_p$. To prove this we first show that the identity

$$\mathbf{T} = \mathbf{T}\mathbf{A}^H\big(\mathbf{A}\mathbf{A}^H\big)^{-1}\mathbf{A} \tag{3.212}$$

with

$$\mathbf{T} = \mathbf{I} - \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{P} \tag{3.213}$$

holds. For that we utilize the matrix $[\mathbf{A}^H \;\; \mathbf{N}]$. Since $\mathbf{A}\mathbf{N} = \mathbf{0}$, the column spaces of $\mathbf{A}^H$ and $\mathbf{N}$ are orthogonal to each other such that $[\mathbf{A}^H \;\; \mathbf{N}]$ is invertible. Multiplying (3.212) with $[\mathbf{A}^H \;\; \mathbf{N}]$ from the right results in $[\mathbf{T}\mathbf{A}^H \;\; \mathbf{0}] = [\mathbf{T}\mathbf{A}^H \;\; \mathbf{0}]$. Since this equation is true and $[\mathbf{A}^H \;\; \mathbf{N}]$ is invertible, (3.212) is also true. Now replacing $\mathbf{T} = \mathbf{I} - \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{P}$ in the second line of (3.198) by the right hand side of (3.212) gives

$$\hat{\mathbf{x}}_{\mathrm{CB}} = \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} + \mathbf{T}\mathbf{A}^H\big(\mathbf{A}\mathbf{A}^H\big)^{-1}\mathbf{A}\mathbf{x}_p \tag{3.214}$$

$$= \mathbf{N}\big(\mathbf{N}^H\mathbf{P}\mathbf{N}\big)^{-1}\mathbf{N}^H\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} + \mathbf{T}\mathbf{A}^H\big(\mathbf{A}\mathbf{A}^H\big)^{-1}\mathbf{b}. \tag{3.215}$$

That finally means that using any particular $\mathbf{x}_p$ in (3.198) yields the same result as using the minimum norm solution $\mathbf{x}_p = \mathbf{A}^H\big(\mathbf{A}\mathbf{A}^H\big)^{-1}\mathbf{b}$. $\square$

Another important note is that $N_{\mathbf{y}} > N_{\mathbf{x}}$ is not required for the application of (3.199), which is in contrast to the constrained LS estimator in (3.29). In fact, the constrained BLUE in (3.199) only requires full column rank of $\mathbf{HN}$ in order for $\mathbf{N}^H\mathbf{PN}$ to be invertible. This implies that $N_{\mathbf{y}} \geq N_0$, but $N_{\mathbf{y}} \leq N_{\mathbf{x}}$ is allowed.

For the case that $N_{\mathbf{y}} > N_{\mathbf{x}}$, $\mathbf{H}$ full column rank, and $\mathbf{C}_{\mathbf{nn}}$ is invertible (as originally assumed) which implies that $\mathbf{P}$ is invertible, the expression for the constrained BLUE in (3.199) can be simplified.

Simplification for invertible $\mathbf{P}$

Note that the expression of the constrained BLUE in (3.199) requires the calculation of a basis of the nullspace of the matrix $\mathbf{A}$. We will now derive an expression of the constrained BLUE that does not require this nullspace evaluation, but which requires $N_{\mathbf{y}} > N_{\mathbf{x}}$. With the assumptions of full column rank $\mathbf{H}$ and invertible $\mathbf{C}_{\mathbf{nn}}$ (as originally assumed) $\mathbf{P}$ is invertible, and the following identity holds:

$$\mathbf{N}\big(\mathbf{N}^H\mathbf{PN}\big)^{-1}\mathbf{N}^H = \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{A}^H\big(\mathbf{A}\mathbf{P}^{-1}\mathbf{A}^H\big)^{-1}\mathbf{A}\mathbf{P}^{-1}. \tag{3.216}$$

This identity can be proven the following way. The i^{th} column of $\mathbf{N}$ is denoted as $\mathbf{x}_i$ according to (3.183). Furthermore, the i^{th} column of $\mathbf{A}^H$ is denoted as $\mathbf{a}_i$. We first show that the vectors $\mathbf{P}^{-1}\mathbf{a}_1, \ldots, \mathbf{P}^{-1}\mathbf{a}_{N_{\mathbf{b}}}, \mathbf{x}_1, \ldots, \mathbf{x}_{N_0}$ are linearly independent: Fix $c_1, \ldots, c_{N_{\mathbf{b}}}, d_i, \ldots, d_{N_0} \in \mathbb{C}$ such that

$$\sum_{i=1}^{N_{\mathbf{b}}} c_i\mathbf{P}^{-1}\mathbf{a}_i + \sum_{j=1}^{N_0} d_i\mathbf{x}_i = \mathbf{0}. \tag{3.217}$$

For $\mathbf{u} = \sum_{i=1}^{N_{\mathbf{b}}} c_i\mathbf{a}_i$ and $\mathbf{v} = \sum_{j=1}^{N_0} d_i\mathbf{x}_i$ we have $\mathbf{P}^{-1}\mathbf{u} + \mathbf{v} = \mathbf{0}^{N_{\mathbf{x}}\times 1}$. Left multiplication by $\mathbf{u}^H$ yields $\mathbf{u}^H\mathbf{P}^{-1}\mathbf{u} = 0$ since $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. Since $\mathbf{P}^{-1}$ is invertible, we have that $\mathbf{u} = \mathbf{0}$. By the linearly independence of all $\mathbf{a}_i$, all c_i are 0. By (3.217), all d_j are 0. Thus the only solution of (3.217) is $c_i = d_j = 0$ for all i, j, or in other words $\mathbf{P}^{-1}\mathbf{a}_1, \ldots, \mathbf{P}^{-1}\mathbf{a}_{N_{\mathbf{b}}}, \mathbf{x}_1, \ldots, \mathbf{x}_{N_0}$ are linearly independent. Hence, the square matrix $[\mathbf{P}^{-1}\mathbf{A}^H \;\; \mathbf{N}]$ is invertible. Furthermore, the matrix $\mathbf{B} = [\mathbf{A}^H \;\; \mathbf{PN}]$ is invertible. Right multiplying (3.216) by $\mathbf{B}$ yields $[\mathbf{0} \;\; \mathbf{N}] = [\mathbf{P}^{-1}\mathbf{A}^H \;\; \mathbf{N}] - [\mathbf{P}^{-1}\mathbf{A}^H \;\; \mathbf{0}]$. Since this equation is true and $\mathbf{B}$ is invertible, (3.216) is also true. $\square$

Inserting (3.216) into (3.199) finally yields

$$\hat{\mathbf{x}}_{\text{CB}} = \Big(\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}^H\big(\mathbf{A}\mathbf{P}^{-1}\mathbf{A}^H\big)^{-1}\mathbf{A}\Big)\mathbf{P}^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} + \mathbf{P}^{-1}\mathbf{A}^H\big(\mathbf{A}\mathbf{P}^{-1}\mathbf{A}^H\big)^{-1}\mathbf{b}. \tag{3.218}$$

For the constrained BLUE in (3.218) one can easily show that the covariance matrix is

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\text{CB}} = \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{A}^H\big(\mathbf{A}\mathbf{P}^{-1}\mathbf{A}^H\big)^{-1}\mathbf{A}\mathbf{P}^{-1}. \tag{3.219}$$

The expression for the constrained BLUE in (3.218) has the advantage that the nullspace of $\mathbf{A}$ is not required. Furthermore, comparing the constrained LS estimator in (3.29) with the constrained BLUE in (3.218) reveals that they are connected in a very similar way as it is the case for the LS estimator in (3.5) and the BLUE in (3.48). Finally, we end up with the following

Result 3.4 (Constrained BLUE)

Consider the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ is the measurement vector, $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a known measurement matrix with $N_\mathbf{y} > N_\mathbf{x}$ and full column rank, and $\mathbf{n} \in \mathbb{C}^{N_\mathbf{y}}$ is a zero mean random noise vector with known invertible covariance matrix $\mathbf{C_{nn}}$. If $\mathbf{x}$ fulfills the linear constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$ with full row rank $\mathbf{A} \in \mathbb{C}^{N_\mathbf{b} \times N_\mathbf{x}}$, $\mathbf{b} \in \mathbb{C}^{N_\mathbf{b}}$, $N_\mathbf{b} < N_\mathbf{x}$, then the constrained BLUE minimizing the variances of the elements of $\hat{\mathbf{x}}_{\mathrm{CB}}$ such that $\hat{\mathbf{x}}_{\mathrm{CB}}$ fulfills $\mathbf{A}\hat{\mathbf{x}}_{\mathrm{CB}} = \mathbf{b}$ is given by (3.218). Its covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{CB}}$ is given by (3.219).

If $N_\mathbf{y} > N_\mathbf{x}$ does not hold, then let $\mathbf{N} \in \mathbb{C}^{N_\mathbf{x} \times N_0}$ be the matrix built by linearly independent (column) basis vectors that span the nullspace of $\mathbf{A}$. If $\mathbf{HN}$ has full column rank (implying $N_\mathbf{y} \geq N_0$), then the constrained BLUE for $\mathbf{x}$ fulfilling $\mathbf{A}\hat{\mathbf{x}}_{\mathrm{CB}} = \mathbf{b}$ is given by (3.199). Its covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{CB}}$ is given by (3.211).

3.3.2 Constrained BWLUE

The extension of Result 3.4 for the case of improper noise is presented in the following. These derivations will lead to the constrained BWLUE. For that, the linear model in (3.1) shall hold and the augmented noise covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$ shall be invertible. A discussion about the relation between $N_\mathbf{y}$ and $N_\mathbf{x}$ will be presented. Again, $\mathbf{A}\mathbf{x} = \mathbf{b}$ shall hold. We assume the estimator to be widely affine and of the form

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{F}\mathbf{y} + \mathbf{G}\mathbf{y}^* + \mathbf{f} && (3.220) \\ &= \underbrace{\begin{bmatrix} \mathbf{F} & \mathbf{G} \end{bmatrix}}_{\mathbf{E}} \underline{\mathbf{y}} + \mathbf{f} && (3.221) \\ &= \mathbf{E}\underline{\mathbf{y}} + \mathbf{f}. && (3.222) \end{aligned}$$

The goal is to find the estimator matrix $\mathbf{E} \in \mathbb{C}^{N_\mathbf{x} \times 2N_\mathbf{y}}$ and the vector $\mathbf{f} \in \mathbb{C}^{N_\mathbf{x}}$. Here, we consider the estimator in augmented notation such that

$$\begin{aligned} \underline{\hat{\mathbf{x}}} &= \underbrace{\begin{bmatrix} \mathbf{F} & \mathbf{G} \\ \mathbf{G}^* & \mathbf{F}^* \end{bmatrix}}_{\underline{\mathbf{E}}} \underline{\mathbf{y}} + \underbrace{\begin{bmatrix} \mathbf{f} \\ \mathbf{f}^* \end{bmatrix}}_{\underline{\mathbf{f}}} && (3.223) \\ &= \underline{\mathbf{E}}\underline{\mathbf{y}} + \underline{\mathbf{f}}. && (3.224) \end{aligned}$$

For that, the unbiased constraint enforces

$$
\begin{aligned}
E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}] &= E_{\mathbf{y}}[\underline{\mathbf{E}}\,\underline{\mathbf{y}} + \underline{\mathbf{f}}] && (3.225)\\
&= E_{\mathbf{n}}[\underline{\mathbf{E}}(\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{n}}) + \underline{\mathbf{f}}] && (3.226)\\
&= \underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{f}} \stackrel{!}{=} \underline{\mathbf{x}}. && (3.227)
\end{aligned}
$$

By using the notation in (3.174), and by denoting the i^{th} row of $\underline{\mathbf{E}}$ as $\mathbf{e}_i^H$, the unbiased constraint for $\mathbf{e}_i^H$ can be extracted from (3.227) and leads to

$$
\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + f_i = x_i \tag{3.228}
$$

for $i = 1, \ldots, N_{\mathbf{x}}$. Moreover, $\hat{x}_i$ can be written as

$$
\hat{x}_i = \mathbf{e}_i^H \underline{\mathbf{y}} + f_i. \tag{3.229}
$$

The cost function to be minimized, which is the variance of $\hat{x}_i$, follows as

$$
\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y}}\left[(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])(\hat{x}_i - E_{\mathbf{y}}[\hat{x}_i])^H\right] && (3.230)\\
&= E_{\mathbf{y}}\left[\left(\mathbf{e}_i^H \underline{\mathbf{y}} + f_i - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] - f_i\right)\left(\mathbf{e}_i^H \underline{\mathbf{y}} + f_i - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] - f_i\right)^H\right] && (3.231)\\
&= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H (\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{n}}) - \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}}\right)\left(\mathbf{e}_i^H (\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{n}}) - \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}}\right)^H\right] && (3.232)\\
&= E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H \underline{\mathbf{n}}\right)\left(\mathbf{e}_i^H \underline{\mathbf{n}}\right)^H\right] = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i. && (3.233)
\end{aligned}
$$

In complete analogy to the linear case in Section 3.3.1 the constraints (3.169) need to be converted into constraints in $\mathbf{e}_i$ and f_i for $i = 1, \ldots, N_{\mathbf{x}}$. This is done in the following.

Since $\mathbf{A}\mathbf{x} = \mathbf{b}$ holds, the investigations in (3.181)–(3.183) remain valid. With the notation

$$
\underline{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0}^{N_{\mathbf{b}} \times N_{\mathbf{x}}} \\ \mathbf{0}^{N_{\mathbf{b}} \times N_{\mathbf{x}}} & \mathbf{A}^* \end{bmatrix}, \qquad \underline{\mathbf{N}} = \begin{bmatrix} \mathbf{N} & \mathbf{0}^{N_{\mathbf{x}} \times N_0} \\ \mathbf{0}^{N_{\mathbf{x}} \times N_0} & \mathbf{N}^* \end{bmatrix}, \tag{3.234}
$$

$$
\underline{\mathbf{x}}_p = \begin{bmatrix} \mathbf{x}_p \\ \mathbf{x}_p^* \end{bmatrix}, \qquad \underline{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^* \end{bmatrix}, \qquad \underline{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ \mathbf{b}^* \end{bmatrix}, \tag{3.235}
$$

the augmented notation of (3.182) is given by

$$
\underline{\mathbf{x}} = \underline{\mathbf{x}}_p + \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}}. \tag{3.236}
$$

Furthermore, it holds that $\underline{\mathbf{A}}\,\underline{\mathbf{N}} = \mathbf{0}^{2N_{\mathbf{b}} \times 2N_0}$. Inserting (3.236) into (3.227) results in

$$
\begin{aligned}
E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}] = \underline{\mathbf{E}}\,\underline{\mathbf{H}}\big(\underline{\mathbf{x}}_p + \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}}\big) + \underline{\mathbf{f}} &\stackrel{!}{=} \underline{\mathbf{x}}_p + \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}} && (3.237)\\
\Leftrightarrow (\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{N}} - \underline{\mathbf{N}})\underline{\boldsymbol{\alpha}} + (\underline{\mathbf{E}}\,\underline{\mathbf{H}} - \mathbf{I})\underline{\mathbf{x}}_p + \underline{\mathbf{f}} &\stackrel{!}{=} \mathbf{0}. && (3.238)
\end{aligned}
$$

To fulfill this equation for every possible realization of the unknown vector $\boldsymbol{\alpha}$, we deduce the following two constraints for $\underline{\mathbf{E}}$ and $\underline{\mathbf{f}}$:

$$
\begin{aligned}
\underline{\mathbf{E}}\,\underline{\mathbf{H}}\,\underline{\mathbf{N}} &= \underline{\mathbf{N}} && (3.239)\\
\underline{\mathbf{f}} &= (\mathbf{I} - \underline{\mathbf{E}}\,\underline{\mathbf{H}})\underline{\mathbf{x}}_p. && (3.240)
\end{aligned}
$$

With $\mathbf{n}_i^H$ denoting the i^{th} row of $\mathbf{N}$, the constraint for $\mathbf{e}_i^H$, $1 \leq i \leq N_{\mathbf{x}}$, can be extracted from (3.239) and leads to

$$\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{N}} = \begin{bmatrix} \mathbf{n}_i^H & \mathbf{0}^{1\times N_0} \end{bmatrix}. \tag{3.241}$$

All the constraints are now converted into constraints in $\mathbf{e}_i$ and f_i, such that we are finally able to formulate the constrained optimization problem for $\mathbf{e}_i$:

$$\mathbf{e}_{\mathrm{CBW},i} = \arg\min_{\mathbf{e}_i} \; \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{N}} = \begin{bmatrix} \mathbf{n}_i^H & \mathbf{0}^{1\times N_0} \end{bmatrix}. \tag{3.242}$$

We now solve this constrained optimization problem using the Lagrangian multiplier method. The Lagrangian cost function for this problem is given by

$$\mathcal{L}(\mathbf{e}_i) = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i + \boldsymbol{\lambda}^H \left(\underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \mathbf{e}_i - \begin{bmatrix} \mathbf{n}_i \\ \mathbf{0}^{N_0 \times 1} \end{bmatrix} \right) + \boldsymbol{\lambda}^T \left(\underline{\mathbf{N}}^T \underline{\mathbf{H}}^T \mathbf{e}_i^* - \begin{bmatrix} \mathbf{n}_i^* \\ \mathbf{0}^{N_0 \times 1} \end{bmatrix} \right). \tag{3.243}$$

The Wirtinger derivative with respect to $\mathbf{e}_i$ produces

$$\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} + \boldsymbol{\lambda}^H \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H. \tag{3.244}$$

Setting (3.244) equal to zero results in

$$\mathbf{e}_{\mathrm{CBW},i}^H = -\boldsymbol{\lambda}^H \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.245}$$

Inserting (3.245) into the constraint in (3.242) produces

$$-\boldsymbol{\lambda}^H \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{N}} = \begin{bmatrix} \mathbf{n}_i^H & \mathbf{0}^{1\times N_0} \end{bmatrix} \tag{3.246}$$

$$-\boldsymbol{\lambda}^H = \begin{bmatrix} \mathbf{n}_i^H & \mathbf{0}^{1\times N_0} \end{bmatrix} \left(\underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{N}} \right)^{-1}. \tag{3.247}$$

Reinserting this result into the expression for $\mathbf{e}_i^H$ in (3.245) yields

$$\mathbf{e}_{\mathrm{CBW},i}^H = \begin{bmatrix} \mathbf{n}_i^H & \mathbf{0}^{1\times N_0} \end{bmatrix} \left(\underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{N}} \right)^{-1} \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.248}$$

Since $\mathbf{n}_i^H$ is the only term in (3.248) that depends on the index i, the expression for the estimator matrix is given by

$$\mathbf{E}_{\mathrm{CBW}} = \begin{bmatrix} \mathbf{N} & \mathbf{0}^{N_{\mathbf{x}}\times N_0} \end{bmatrix} \left(\underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{N}} \right)^{-1} \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.249}$$

The augmented notation of this result can easily be identified as

$$\underline{\mathbf{E}}_{\mathrm{CBW}} = \underline{\mathbf{N}} \left(\underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{N}} \right)^{-1} \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.250}$$

In the following, we denote

$$\underline{\mathbf{P}} = \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}} \tag{3.251}$$

such that $\underline{\mathbf{E}}_{\mathrm{CBW}}$ reads as

$$\underline{\mathbf{E}}_{\mathrm{CBW}} = \underline{\mathbf{N}} \left(\underline{\mathbf{N}}^H \underline{\mathbf{P}}\,\underline{\mathbf{N}} \right)^{-1} \underline{\mathbf{N}}^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.252}$$

Inserting (3.240) and (3.252) into (3.224) finally leads to the constrained BWLUE in the form of

$$\begin{aligned}
\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}} &= \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{y}} + (\mathbf{I} - \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}})\underline{\mathbf{x}}_p && (3.253)\\
&= \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} + \Big(\mathbf{I} - \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{P}}\Big)\underline{\mathbf{x}}_p && (3.254)\\
&= \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\big(\underline{\mathbf{y}} - \underline{\mathbf{H}}\,\underline{\mathbf{x}}_p\big) + \underline{\mathbf{x}}_p. && (3.255)
\end{aligned}$$

This estimator is unbiased, which can be shown by incorporating (3.236) and (3.239) into (3.253)

$$\begin{aligned}
E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}}] &= \underline{\mathbf{E}}_{\mathrm{CBW}}E_{\mathbf{y}}[\underline{\mathbf{y}}] + (\mathbf{I} - \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}})\underline{\mathbf{x}}_p && (3.256)\\
&= \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{x}} + (\mathbf{I} - \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}})(\underline{\mathbf{x}} - \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}}) && (3.257)\\
&= \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{x}} - \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{x}} - \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}} + \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}} && (3.258)\\
&= \underline{\mathbf{x}} - \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}} + \underbrace{\underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{N}}}_{\underline{\mathbf{N}}}\underline{\boldsymbol{\alpha}} && (3.259)\\
&= \underline{\mathbf{x}}. && (3.260)
\end{aligned}$$

Following similar arguments, it holds that

$$\begin{aligned}
\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}} - E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}}] &= \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{x}} + \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{n}} + \underline{\mathbf{x}} - \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{x}} - \underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}} + \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{H}}\,\underline{\mathbf{N}}\,\underline{\boldsymbol{\alpha}} - \underline{\mathbf{x}} && (3.261)\\
&= \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{n}}. && (3.262)
\end{aligned}$$

With that, the augmented covariance matrix of $\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}}$ can be derived as

$$\begin{aligned}
\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{CBW}} &= E_{\mathbf{y}}\Big[(\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}} - E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}}])(\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}} - E_{\mathbf{y}}[\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}}])^H\Big] && (3.263)\\
&= \underline{\mathbf{E}}_{\mathrm{CBW}}E_{\mathbf{n}}\big[\underline{\mathbf{n}}\,\underline{\mathbf{n}}^H\big]\underline{\mathbf{E}}_{\mathrm{CBW}}^H && (3.264)\\
&= \underline{\mathbf{E}}_{\mathrm{CBW}}\underline{\mathbf{C}}_{\mathbf{nn}}\underline{\mathbf{E}}_{\mathrm{CBW}}^H && (3.265)\\
&= \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{C}}_{\mathbf{nn}}\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\,\underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H && (3.266)\\
&= \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{H}}^H\underline{\mathbf{P}}\,\underline{\mathbf{H}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H. && (3.267)
\end{aligned}$$

In complete analogy to the linear case, one can show that $\underline{\hat{\mathbf{x}}}_{\mathrm{CBW}}$ in (3.255) actually is independent of the concrete choice of $\mathbf{x}_p$ as long as $\underline{\mathbf{A}}\,\underline{\mathbf{x}}_p = \underline{\mathbf{b}}$. To prove this we first show that the identity

$$\underline{\mathbf{T}} = \underline{\mathbf{T}}\,\underline{\mathbf{A}}^H\big(\underline{\mathbf{A}}\,\underline{\mathbf{A}}^H\big)^{-1}\underline{\mathbf{A}}, \tag{3.268}$$

with

$$\underline{\mathbf{T}} = \mathbf{I} - \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{P}}, \tag{3.269}$$

holds. For that we utilize the matrix $[\underline{\mathbf{A}}^H \;\; \underline{\mathbf{N}}]$. Since $\underline{\mathbf{A}}\,\underline{\mathbf{N}} = \mathbf{0}$, the column spaces of $\underline{\mathbf{A}}^H$ and $\underline{\mathbf{N}}$ are orthogonal to each such that $[\underline{\mathbf{A}}^H \;\; \underline{\mathbf{N}}]$ is invertible. Multiplying (3.268) with $[\underline{\mathbf{A}}^H \quad \underline{\mathbf{N}}]$ from the right results in $[\underline{\mathbf{T}}\,\underline{\mathbf{A}}^H \quad \mathbf{0}] = [\underline{\mathbf{T}}\,\underline{\mathbf{A}}^H \quad \mathbf{0}]$. Since this equation is true and $[\underline{\mathbf{A}}^H \quad \underline{\mathbf{N}}]$ is invertible, (3.268) is also true. $\square$

Now replacing $\underline{\mathbf{T}} = \mathbf{I} - \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{P}}$ in (3.254) by the right hand side of (3.268) gives

$$\underline{\hat{\mathbf{x}}}_{\text{CBW}} = \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} + \underline{\mathbf{T}}\underline{\mathbf{A}}^H\big(\underline{\mathbf{A}}\,\underline{\mathbf{A}}^H\big)^{-1}\underline{\mathbf{A}}\mathbf{x}_p \tag{3.270}$$

$$= \underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} + \underline{\mathbf{T}}\underline{\mathbf{A}}^H\big(\underline{\mathbf{A}}\,\underline{\mathbf{A}}^H\big)^{-1}\underline{\mathbf{b}}. \tag{3.271}$$

That finally means that using any particular $\mathbf{x}_p$ in (3.254) yields the same result as using the minimum norm solution $\mathbf{x}_p = \mathbf{A}^H\big(\mathbf{A}\mathbf{A}^H\big)^{-1}\mathbf{b}$.

Another important note is that $N_\mathbf{y} > N_\mathbf{x}$ is not required for the application of (3.255). In fact, the constrained BWLUE in (3.255) only requires full column rank of $\underline{\mathbf{HN}}$ in order for $\underline{\mathbf{N}}^H\underline{\mathbf{PN}}$ to be invertible. This implies that $N_\mathbf{y} \geq N_0$, but $N_\mathbf{y} \leq N_\mathbf{x}$ is allowed.

For the case that $N_\mathbf{y} > N_\mathbf{x}$, $\mathbf{H}$ full column rank, and $\underline{\mathbf{C}}_{\mathbf{nn}}$ is invertible (as originally assumed) which implies that $\mathbf{P}$ is invertible, the expression for the constrained BWLUE in (3.255) can be simplified.

Simplification for invertible $\underline{\mathbf{P}}$

Note that the expression of the constrained BWLUE in (3.255) requires the evaluation of the nullspace of the matrix $\mathbf{A}$. We will now derive an expression of the constrained BWLUE that does not require this nullspace evaluation, but which requires $N_\mathbf{y} > N_\mathbf{x}$ and $\underline{\mathbf{P}}$ being invertible. For that we utilize the identity

$$\underline{\mathbf{N}}\big(\underline{\mathbf{N}}^H\underline{\mathbf{P}}\,\underline{\mathbf{N}}\big)^{-1}\underline{\mathbf{N}}^H = \underline{\mathbf{P}}^{-1} - \underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\big(\underline{\mathbf{A}}\,\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\big)^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{P}}^{-1}. \tag{3.272}$$

This identity can be proven the following way. Let the i^{th} column of $\underline{\mathbf{N}}$ be denoted as $\widetilde{\mathbf{n}}_i$. Furthermore, the i^{th} column of $\underline{\mathbf{A}}^H$ is denoted as $\widetilde{\mathbf{a}}_i$. We first show that the vectors $\underline{\mathbf{P}}^{-1}\widetilde{\mathbf{a}}_1, \ldots, \underline{\mathbf{P}}^{-1}\widetilde{\mathbf{a}}_{2N_\mathbf{b}}, \widetilde{\mathbf{n}}_1, \ldots, \widetilde{\mathbf{n}}_{2N_0}$ are linearly independent:
Fix $c_1, \ldots, c_{2N_\mathbf{x}}, d_i, \ldots, d_{2N_0} \in \mathbb{C}$ such that

$$\sum_{i=1}^{2N_\mathbf{b}} c_i\underline{\mathbf{P}}^{-1}\widetilde{\mathbf{a}}_i + \sum_{j=1}^{N_0} d_i\widetilde{\mathbf{n}}_i = \mathbf{0}. \tag{3.273}$$

For $\mathbf{u} = \sum_{i=1}^{N_\mathbf{b}} c_i\widetilde{\mathbf{a}}_i$ and $\mathbf{v} = \sum_{j=1}^{N_0} d_i\widetilde{\mathbf{n}}_i$ we have $\underline{\mathbf{P}}^{-1}\mathbf{u} + \mathbf{v} = \mathbf{0}^{N_\mathbf{x}\times 1}$. Left multiplication by $\mathbf{u}^H$ yields $\mathbf{u}^H\underline{\mathbf{P}}^{-1}\mathbf{u} = -\mathbf{u}^H\mathbf{v} = 0$ since $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. Since $\underline{\mathbf{P}}^{-1}$ is invertible, we have that $\mathbf{u} = \mathbf{0}$. By the linearly independence of all $\widetilde{\mathbf{a}}_i$, all c_i are 0. By (3.273), all d_j are 0. Thus the only solution of (3.273) is $c_i = d_j = 0$ for all i, j, or in other words $\underline{\mathbf{P}}^{-1}\widetilde{\mathbf{a}}_1, \ldots, \underline{\mathbf{P}}^{-1}\widetilde{\mathbf{a}}_{2N_\mathbf{b}}, \widetilde{\mathbf{n}}_1, \ldots, \widetilde{\mathbf{n}}_{2N_0}$ are linearly independent. Hence, the square matrix $[\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H \quad \underline{\mathbf{N}}]$ is invertible. Furthermore, the matrix $\mathbf{B} = [\underline{\mathbf{A}}^H \quad \underline{\mathbf{P}}\,\underline{\mathbf{N}}]$ is invertible. Right multiplying (3.272) by $\mathbf{B}$ yields $[\mathbf{0} \quad \underline{\mathbf{N}}] = [\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H \quad \underline{\mathbf{N}}] - [\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H \quad \mathbf{0}]$. Since this equation is true and $\mathbf{B}$ is invertible, (3.272) is also true. $\square$

Inserting (3.272) into (3.255) finally yields

$$\underline{\hat{\mathbf{x}}}_{\text{CBW}} = \left(\mathbf{I} - \underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\left(\underline{\mathbf{A}}\,\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\right)^{-1}\underline{\mathbf{A}}\right)\underline{\mathbf{P}}^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} + \underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\left(\underline{\mathbf{A}}\,\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\right)^{-1}\underline{\mathbf{b}}. \tag{3.274}$$

For the constrained BWLUE in (3.274) one can easily show that the augmented covariance matrix is

$$\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\text{CBW}} = \underline{\mathbf{P}}^{-1} - \underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\left(\underline{\mathbf{A}}\,\underline{\mathbf{P}}^{-1}\underline{\mathbf{A}}^H\right)^{-1}\underline{\mathbf{A}}\,\underline{\mathbf{P}}^{-1}. \tag{3.275}$$

The expression for the constrained BWLUE in (3.274) has the huge advantage that the nullspace of $\mathbf{A}$ is not required.

All findings for the constrained BWLUE are summarized in

Result 3.5 (Constrained BWLUE)

Consider the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ is the measurement vector, $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a known measurement matrix with $N_\mathbf{y} > N_\mathbf{x}$ and full column rank, and $\mathbf{n} \in \mathbb{C}^{N_\mathbf{y}}$ is a zero mean random noise vector with known invertible augmented covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$. If $\mathbf{x}$ fulfills the linear constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$ with full row rank $\mathbf{A} \in \mathbb{C}^{N_\mathbf{b} \times N_\mathbf{x}}$, $\mathbf{b} \in \mathbb{C}^{N_\mathbf{b}}$, $N_\mathbf{b} < N_\mathbf{x}$, then the constrained BWLUE minimizing the variances of the elements of $\hat{\mathbf{x}}_{\text{CBW}}$ such that $\hat{\mathbf{x}}_{\text{CBW}}$ fulfills $\mathbf{A}\hat{\mathbf{x}}_{\text{CBW}} = \mathbf{b}$ is given in augmented notation by (3.274). Its covariance matrix $\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\text{CBW}}$ is given by (3.275).

If $N_\mathbf{y} > N_\mathbf{x}$ does not hold, then let $\mathbf{N} \in \mathbb{C}^{N_\mathbf{x} \times N_0}$ be the matrix built by linearly independent (column) basis vectors that span the nullspace of $\mathbf{A}$. If $\underline{\mathbf{H}}\,\underline{\mathbf{N}}$ has full column rank (implying $N_\mathbf{y} \geq N_0$), then the constrained BWLUE for $\mathbf{x}$ fulfilling $\mathbf{A}\hat{\mathbf{x}}_{\text{CBW}} = \mathbf{b}$ is given in augmented notation by (3.255). Its augmented covariance matrix $\underline{\mathbf{C}}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\text{CBW}}$ is given by (3.267).

The derived estimators in Result 3.4 and Result 3.5 will now be compared to competing estimators in two simulation examples.

Example 3.1 (Estimation of an Impulse Response Whose Samples Sum up to Zero)

We assume $\mathbf{x} \in \mathbb{C}^5$ to be the discrete-time impulse response of an unknown system. Additionally, we know that the system is unable to transmit any DC signals. Hence, the sum of all elements of $\mathbf{x}$ must be zero. This can be described by a linear constraint $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} = \mathbf{1}^{1\times 5}$ and $\mathbf{b} = 0$. The measurement vector $\mathbf{y} \in \mathbb{C}^{10}$ shall contain noisy measurements of an input signal $\mathbf{u} \in \mathbb{C}^6$ convolved with the impulse response $\mathbf{x}$. Thus, $\mathbf{H} \in \mathbb{C}^{10\times 5}$ is a convolution matrix built from the vector $\mathbf{u}$. This vector was randomly drawn for every simulation run from a standard proper Gaussian distribution. The noise $\mathbf{n}$ in (3.346) is chosen by [32]

$$\mathbf{n} = \sqrt{1-\rho^2}\mathbf{n}_r + j\rho\mathbf{n}_i, \tag{3.276}$$

where $\mathbf{n}_r$ and $\mathbf{n}_i$ are uncorrelated real-valued zero mean Gaussian random vectors of length $N_\mathbf{y}$ and with unit variance, and $\rho \in [0,1]$. With that choice, the noise power remains unaffected while the improperness of the noise can be adjusted by appropriately choosing ρ. The noise is proper for $\rho = 1/\sqrt{2}$. Since all elements of $\mathbf{n}$ have the same variance, the MSE performance of the ordinary BLUE coincides with that of the ordinary LS estimator. Also, the MSE performance of the constrained BLUE coincides with that of the constrained LS estimator. Hence, the following estimators are considered:

1. The ordinary LS estimator in (3.5), denoted as $\hat{\underline{\mathbf{x}}}_{\text{OLS}}$,
2. the intuitive estimator resulting from subtracting the mean value from the estimates of the ordinary LS estimator
$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_{\text{OLS}} - \text{mean}(\hat{\mathbf{x}}_{\text{OLS}})\mathbf{1}^{N_\mathbf{x}\times 1}, \tag{3.277}$$
3. the constrained LS estimator in (3.29),
4. the ordinary BWLUE in (3.63), denoted as $\hat{\underline{\mathbf{x}}}_{\text{OBW}}$,
5. the intuitive estimator resulting from subtracting the mean value from the estimates of the ordinary BWLUE
$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_{\text{OBW}} - \text{mean}(\hat{\mathbf{x}}_{\text{OBW}})\mathbf{1}^{N_\mathbf{x}\times 1}, \tag{3.278}$$
where $\hat{\mathbf{x}}_{\text{OBW}}$ is the upper half of the augmented vector estimate $\hat{\underline{\mathbf{x}}}_{\text{OBW}}$,
6. the constrained BWLUE from Result 3.5.

The resulting average MSEs (averaged over the elements of $\mathbf{x}$) plotted over ρ are presented in Fig. 3.1. The estimators that do not incorporate the improperness of the noise show a performance that is independent of ρ. The ordinary LS estimator performs worst for all values of ρ. The estimation accuracy can be significantly increased by using the intuitive LS estimator in (3.277). An even better estimation accuracy is achieved by the constrained LS estimator. Similar performance gains are obtained for the widely linear estimators based on the BWLUE, however, their performance strongly depends on ρ since they incorporate information about the improperness of the noise.

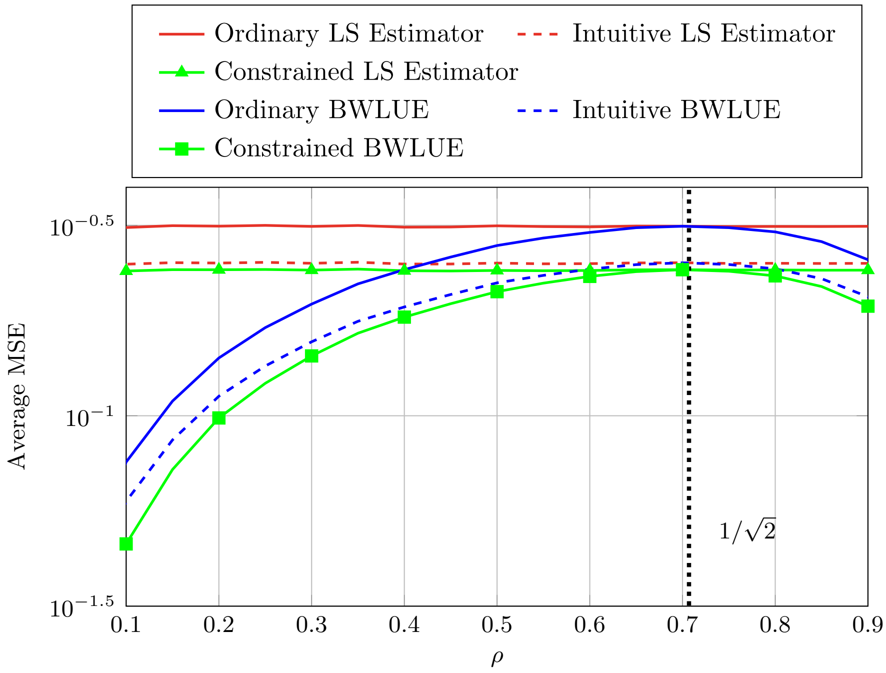

Figure 3.1: Average MSEs of the estimated impulse responses for various estimators. The vertical black line marks the value of $\rho = 1/\sqrt{2}$ where the noise is proper.

Interestingly, if the measurement matrix would fulfill $\mathbf{H}^H\mathbf{H} = \alpha\mathbf{I}$ with an arbitrary scalar $\alpha > 0$, the intuitive LS estimator in (3.277) and the constrained LS estimator in (3.29) would coincide. This equivalence is analytically proven in Appendix C.

In the next example, $\mathbf{b}$ differs from the zero vector.

Example 3.2 (Estimation of a Parameter Vector that fulfills Additional Constraints)

This toy example is based on the previous one. In fact, the measurement matrix generation and the noise statistics remain unaltered, but we change the constraints that the unknown parameter vector has to fulfill. Here, the linear constraints are $\mathbf{A}\mathbf{x} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 & 0.1 & 0.5 & 1 \\ 0 & 0.5 & 1 & 0.5 & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{3.279}$$

For these settings, the same estimators as in Example 3.1 are compared except for the intuitive ones since they do not make sense in that form in this example.

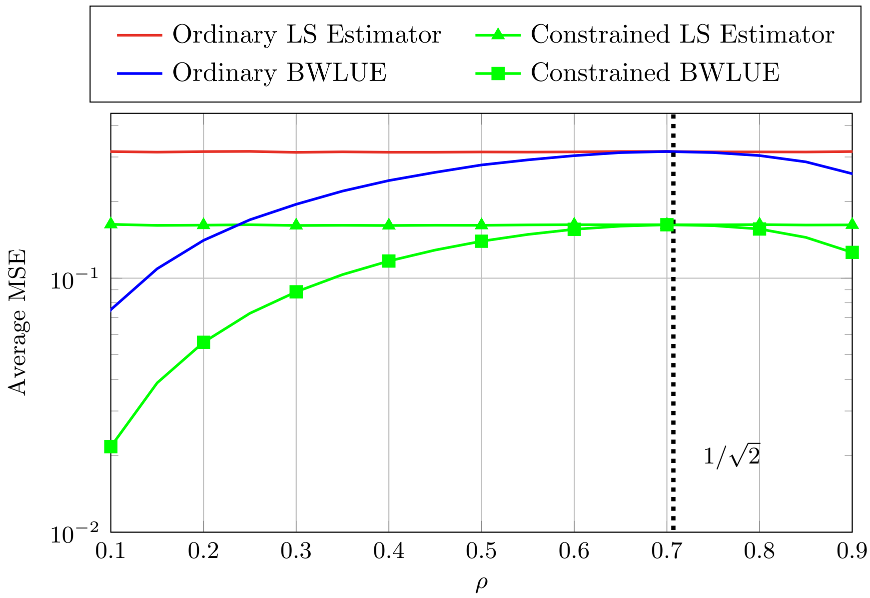

Figure 3.2: Average MSEs of the estimated impulse responses for various estimators. The vertical black line marks the value of $\rho = 1/\sqrt{2}$ where the noise is proper.

The resulting average MSEs (averaged over the elements of $\mathbf{x}$) plotted over ρ are presented in Fig. 3.2. Compared to Example 3.1, the performance gain achievable by incorporating the additional constraints into the estimation process increased. This is due to the increased number of additional scalar valued constraints ($N_{\mathbf{b}} = 2$).

Another type of additional model knowledge is considered in the next section. The derived estimators therein not only significantly outperform well-known estimators but also meaningful intuitive estimators as it will be shown.

3.4 Estimation of Real-Valued Parameter Vectors in Complex-Valued Environments

Consider the linear model in (3.1) with the additional model knowledge that the parameter vector $\mathbf{x}$ is real-valued. The measurement matrix $\mathbf{H}$, the noise $\mathbf{n}$ as well as the measurements $\mathbf{y}$ remain complex-valued. A prominent practical example where such a model appears is the estimation of a real-valued impulse response of a LTI system based on complex-valued noisy measurements of its frequency response. Standard complex-valued estimators such as the LS estimator, the BLUE or the BWLUE result in complex-valued estimates in that case. This fact corresponds to a systematic error. However, by incorporating the fact that $\mathbf{x}$ is real-valued into the derivations, the goal of finding classical estimators that results in real-valued estimates can be achieved.

We assume an underlying linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.280}$$

where, in contrast to (3.1), the parameter vector $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$ is assumed to be real-valued. The measurement vector $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$, the measurement matrix $\mathbf{H} \in \mathbb{C}^{N_{\mathbf{y}} \times N_{\mathbf{x}}}$, and the noise vector $\mathbf{n} \in \mathbb{C}^{N_{\mathbf{y}}}$ remain complex-valued.

In a Bayesian interpretation, the real-valued parameter vector is improper and the application of widely linear estimators is obvious [2,3,22,33–35]. A common widely linear Bayesian estimator is the WLMMSE estimator discussed in Section 4. The WLMMSE estimator incorporates the fact that $\mathbf{x}$ is real-valued and produces real-valued estimates. However, it requires prior knowledge about $\mathbf{x}$ in form of augmented first and second order statistics, which may not always be available. If this is the case, the classical estimators derived in this section may be an optimal choice.

Note that standard classical estimators such as the LS estimator in (3.5), the BLUE in (3.48), or the BWLUE in (3.63) do not result in real-valued estimates for real-valued parameters in general. An exception is the case where $\mathbf{H}$, $\mathbf{x}$, $\mathbf{n} \in \mathbb{R}$. Hence, these classical estimators produce a systematic error when applied to (3.280). A common practical approach to overcome this issue is to take only the real parts of the estimates for further processing. However, this approach is in general not optimal as will be shown shortly. A special case where this practical approach turns out to be optimal is discussed later. Also note that the nomenclature BLUE and BWLUE for the estimators (3.48) and (3.63) applied on the model in (3.280) is misleading since they will no longer be the true best (widely) linear unbiased estimators any more for the particular model. However, for the sake of uniformity, we continue to refer (3.48) and (3.63) as the ordinary BLUE and ordinary BWLUE, respectively.

In the following, we derive a classical estimator termed BWLUE *for real-valued parameter vectors.* This estimator is of widely linear form and incorporates the fact that $\mathbf{x}$ is real-valued in an optimal way. After that, the LS estimator is adapted to the model in (3.280). We derive a widely linear LS estimator that does not utilize any noise statistics, and that incorporates the fact that $\mathbf{x}$ is real-valued in an optimal way. The resulting estimator is termed the widely linear least squares (WLLS) *estimator for real-valued parameter vectors*.

We note that a similar results as presented in this section are already published in [36] by means of a matched filter and for the special case of a scalar parameter. Another similar result can be found in [35] for the special case of diagonal measurement matrices. We also note that the approach in (3.308) should not be confused with beamforming approaches based on the minimum variance distortionless response (MVDR) beamformer [37,38]. There, the utilized correlation matrix contains contributions from all incoming signals and noise. In our approach, no statistics about the signal (the parameter vector in this context) is required.

3.4.1 BWLUE for Real-Valued Parameter Vectors

In this section, we derive the BWLUE for real-valued parameter vectors but complex-valued measurements. Utilizing the notation introduced in (3.51) gives us an expression for a general widely linear estimator for x_i as

$$\hat{x}_i = \mathbf{f}_i^H \mathbf{y} + \mathbf{g}_i^H \mathbf{y}^*. \tag{3.281}$$

In contrast to the ordinary BWLUE in (3.63), the BWLUE for real-valued parameter vectors enforces

$$\mathrm{Im}\{\hat{x}_i\} = 0 \tag{3.282}$$

and

$$E_{\mathbf{y}}[\mathrm{Re}\{\hat{x}_i\}] = E_{\mathbf{y}}[\hat{x}_i] = x_i. \tag{3.283}$$

From (3.282), one can easily show that the choice

$$\mathbf{f}_i^H = \mathbf{g}_i^T \tag{3.284}$$

is necessary and sufficient to make $\hat{x}_i$ real-valued, independent of the actual realization of $\mathbf{y}$. The proof of this statement is provided in Appendix D. Incorporating (3.281) and (3.284) into (3.283) leads to

$$E_{\mathbf{y}}[\hat{x}_i] = E_{\mathbf{y}}\left[\mathbf{f}_i^H \mathbf{y} + \mathbf{f}_i^T \mathbf{y}^*\right] \tag{3.285}$$

$$= \mathbf{f}_i^H \mathbf{H}\mathbf{x} + \mathbf{f}_i^T \mathbf{H}^* \mathbf{x} \tag{3.286}$$

$$= \left(\mathbf{f}_i^H \mathbf{H} + \mathbf{f}_i^T \mathbf{H}^*\right)\mathbf{x}. \tag{3.287}$$

Hence, the unbiased constraint $E[\hat{x}_i] = \hat{x}_i$ is fulfilled for every $\mathbf{x}$ if

$$\mathbf{f}_i^H \mathbf{H} + \mathbf{f}_i^T \mathbf{H}^* = \mathbf{u}_i^T, \tag{3.288}$$

with $\mathbf{u}_i^T$ being a row vector of size $1 \times N_{\mathbf{x}}$ with a '1' at its i^{th} position, and all zeros elsewhere. Comparing the constraints for the BWLUE for real-valued parameter vectors in (3.288) with that for the ordinary BWLUE in (3.53) reveals that the ordinary BWLUE has to fulfill twice as many constraints than the BWLUE for real-valued parameter vectors. On the other hand, the BWLUE for real-valued parameter vectors only has half as many degrees of freedom compared to the ordinary BWLUE due to (3.284).

Combining (3.288) with (3.59) allows to derive the constrained optimization problem

$$\mathbf{f}_{\mathrm{BW},i} = \arg\min_{\mathbf{f}_i} \left(\begin{bmatrix} \mathbf{f}_i^H & \mathbf{f}_i^T \end{bmatrix} \underline{\mathbf{C}}_{\mathbf{nn}} \begin{bmatrix} \mathbf{f}_i \\ \mathbf{f}_i^* \end{bmatrix} \right) \tag{3.289}$$

$$= \arg\min_{\mathbf{f}_i} \left(2\mathbf{f}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{f}_i + \mathbf{f}_i^H \widetilde{\mathbf{C}}_{\mathbf{nn}} \mathbf{f}_i^* + \mathbf{f}_i^T \widetilde{\mathbf{C}}_{\mathbf{nn}}^* \mathbf{f}_i \right) \tag{3.290}$$

$$\text{s.t.} \quad \mathbf{f}_i^H \mathbf{H} + \mathbf{f}_i^T \mathbf{H}^* = \mathbf{u}_i^T. \tag{3.291}$$

This can again be solved by utilizing the Lagrange multiplier method. Since the constraint is real-valued independent of $\mathbf{f}_i$ (see discussion in Section 2.4.2) the Lagrangian cost function follows as

$$\begin{aligned} \mathcal{L}(\mathbf{f}_i) = 2\mathbf{f}_i^H \mathbf{C}_{\mathbf{nn}} \mathbf{f}_i + \mathbf{f}_i^H \widetilde{\mathbf{C}}_{\mathbf{nn}} \mathbf{f}_i^* + \mathbf{f}_i^T \widetilde{\mathbf{C}}_{\mathbf{nn}}^* \mathbf{f}_i \\ + \boldsymbol{\lambda}^T \left(\mathbf{H}^H \mathbf{f}_i + \mathbf{H}^T \mathbf{f}_i^* - \mathbf{u}_i \right). \end{aligned} \tag{3.292}$$

Taking the Wirtinger derivative w.r.t. $\mathbf{f}_i$ results in

$$\frac{\partial \mathcal{L}(\mathbf{f}_i)}{\partial \mathbf{f}_i} = 2\mathbf{f}_i^H \mathbf{C}_{\mathbf{nn}} + 2\mathbf{f}_i^T \widetilde{\mathbf{C}}_{\mathbf{nn}}^* + \boldsymbol{\lambda}^T \mathbf{H}^H, \tag{3.293}$$

where $\boldsymbol{\lambda}$ is real-valued since the constraint is real-valued (cf. Section 2.4.2). Setting (3.293) equal to zero and utilizing $\underline{\mathbf{e}}_{\mathrm{BW},i}^H = \begin{bmatrix} \mathbf{f}_{\mathrm{BW},i}^H & \mathbf{f}_{\mathrm{BW},i}^T \end{bmatrix}$ yields

$$\mathbf{f}_{\mathrm{BW},i}^H \mathbf{C}_{\mathbf{nn}} + \mathbf{f}_{\mathrm{BW},i}^T \widetilde{\mathbf{C}}_{\mathbf{nn}}^* = -\frac{1}{2} \boldsymbol{\lambda}^T \mathbf{H}^H \tag{3.294}$$

$$\underline{\mathbf{e}}_{\mathrm{BW},i}^H \begin{bmatrix} \mathbf{C}_{\mathbf{nn}} \\ \widetilde{\mathbf{C}}_{\mathbf{nn}}^* \end{bmatrix} = -\frac{1}{2} \boldsymbol{\lambda}^T \mathbf{H}^H. \tag{3.295}$$

The complex conjugate of (3.294) can be rewritten in a similar form, producing

$$\underline{\mathbf{e}}_{\mathrm{BW},i}^H \begin{bmatrix} \widetilde{\mathbf{C}}_{\mathbf{nn}} \\ \mathbf{C}_{\mathbf{nn}}^* \end{bmatrix} = -\frac{1}{2} \boldsymbol{\lambda}^T \mathbf{H}^T. \tag{3.296}$$

Combining (3.295) and (3.296) yields

$$\underline{\mathbf{e}}_{\mathrm{BW},i}^H \begin{bmatrix} \mathbf{C}_{\mathbf{nn}} & \widetilde{\mathbf{C}}_{\mathbf{nn}} \\ \widetilde{\mathbf{C}}_{\mathbf{nn}}^* & \mathbf{C}_{\mathbf{nn}}^* \end{bmatrix} = -\frac{1}{2} \boldsymbol{\lambda}^T \underbrace{\begin{bmatrix} \mathbf{H}^H & \mathbf{H}^T \end{bmatrix}}_{\widetilde{\mathbf{H}}^H} \tag{3.297}$$

$$\underline{\mathbf{e}}_{\mathrm{BW},i}^H \underline{\mathbf{C}}_{\mathbf{nn}} = -\frac{1}{2} \boldsymbol{\lambda}^T \widetilde{\mathbf{H}}^H \tag{3.298}$$

$$\underline{\mathbf{e}}_{\mathrm{BW},i}^H = -\frac{1}{2} \boldsymbol{\lambda}^T \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}, \tag{3.299}$$

where

$$\widetilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}^* \end{bmatrix}. \tag{3.300}$$

Inserting (3.299) into the constraint in (3.288) produces

$$\begin{aligned}
\mathbf{f}_{\mathrm{BW},i}^H \mathbf{H} + \mathbf{f}_{\mathrm{BW},i}^T \mathbf{H}^* &= \mathbf{u}_i^T \quad (3.301)\\
\underline{\mathbf{e}}_{\mathrm{BW},i}^H \widetilde{\mathbf{H}} &= \mathbf{u}_i^T \quad (3.302)\\
-\frac{1}{2}\boldsymbol{\lambda}^T \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}} &= \mathbf{u}_i^T \quad (3.303)\\
-\frac{1}{2}\boldsymbol{\lambda}^T &= \mathbf{u}_i^T \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1}. \quad (3.304)
\end{aligned}$$

A reinsertion of (3.304) into (3.299) allows identifying $\underline{\mathbf{e}}_{\mathrm{BW},i}^H$ as

$$\underline{\mathbf{e}}_{\mathrm{BW},i}^H = \mathbf{u}_i^T \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.305}$$

The i^{th} estimate $\hat{x}_i$ follows as

$$\hat{x}_i = \mathbf{f}_{\mathrm{BW},i}^H \mathbf{y} + \mathbf{f}_{\mathrm{BW},i}^T \mathbf{y}^* = \underline{\mathbf{e}}_{\mathrm{BW},i}^H \underline{\mathbf{y}}. \tag{3.306}$$

Since $\mathbf{u}_i^T$ is the only term on the right hand side of (3.305) that depends on the index i, the vector estimate $\hat{\mathbf{x}}$ becomes

$$\begin{aligned}
\hat{\mathbf{x}}_{\mathrm{BW}} &= \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{y}} \quad (3.307)\\
&= \mathbf{E}_{\mathrm{BW}} \underline{\mathbf{y}}, \quad (3.308)
\end{aligned}$$

where

$$\mathbf{E}_{\mathrm{BW}} = \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}. \tag{3.309}$$

We now derive the covariance matrix of $\hat{\mathbf{x}}_{\mathrm{BW}}$. First of all, note that for real-valued $\mathbf{x}$ it holds that

$$\underline{\mathbf{H}}\,\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{x} \end{bmatrix} = \widetilde{\mathbf{H}} \mathbf{x} \tag{3.310}$$

and thus

$$\begin{aligned}
E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{BW}}] &= \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\,\underline{\mathbf{x}} \quad (3.311)\\
&= \mathbf{x}. \quad (3.312)
\end{aligned}$$

Furthermore, it holds that

$$\begin{aligned}
\hat{\mathbf{x}}_{\mathrm{BW}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{BW}}] &= \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{n}} \quad (3.313)\\
&= \mathbf{E}_{\mathrm{BW}} \underline{\mathbf{n}}. \quad (3.314)
\end{aligned}$$

Finally, the covariance matrix of $\hat{\mathbf{x}}_{\mathrm{BW}}$ can be derived as

$$\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{BW}} = E\Big[(\hat{\mathbf{x}}_{\mathrm{BW}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{BW}}])(\hat{\mathbf{x}}_{\mathrm{BW}} - E[\hat{\mathbf{x}}_{\mathrm{BW}}])^H\Big] \tag{3.315}$$

$$= \mathbf{E}_{\mathrm{BW}}\, \underline{\mathbf{C}}_{\mathbf{nn}}\, \mathbf{E}_{\mathrm{BW}}^H \tag{3.316}$$

$$= \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{C}}_{\mathbf{nn}} \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}} \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1} \tag{3.317}$$

$$= \left(\widetilde{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \widetilde{\mathbf{H}}\right)^{-1}. \tag{3.318}$$

The investigations in this section so far are summarized in

Result 3.6 (BWLUE for Real-Valued Parameter Vectors)

If $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$ and $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ are connected via the linear model in (3.280), then the BWLUE for real-valued parameter vectors is given by $\hat{\mathbf{x}}_{\mathrm{BW}} = \mathbf{E}_{\mathrm{BW}}\underline{\mathbf{y}}$, where the estimator matrix $\mathbf{E}_{\mathrm{BW}}$ is defined in (3.309) and $\widetilde{\mathbf{H}}$ is defined in (3.300). This estimator is unbiased in the classical sense, i.e., it fulfills $E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{BW}}] = \mathbf{x}$, and its covariance matrix $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{BW}}$ is given in (3.318).

Several aspects and details of this result are discussed in the following.

Equivalent Real-Valued Model

The complex-valued measurements $\mathbf{y}$ in (3.280) can also be brought into the form of a real composite vector

$$\mathbf{y}_{\mathbb{R}} = \begin{bmatrix} \mathrm{Re}\{\mathbf{y}\} \\ \mathrm{Im}\{\mathbf{y}\} \end{bmatrix}. \tag{3.319}$$

$\mathbf{y}_{\mathbb{R}}$ is connected with the real-valued parameter vector $\mathbf{x}$ via the real composite linear model

$$\mathbf{y}_{\mathbb{R}} = \underbrace{\begin{bmatrix} \mathrm{Re}\{\mathbf{H}\} \\ \mathrm{Im}\{\mathbf{H}\} \end{bmatrix}}_{\mathbf{H}_{\mathbb{R}}} \mathbf{x} + \underbrace{\begin{bmatrix} \mathrm{Re}\{\mathbf{n}\} \\ \mathrm{Im}\{\mathbf{n}\} \end{bmatrix}}_{\mathbf{n}_{\mathbb{R}}} \tag{3.320}$$

$$= \mathbf{H}_{\mathbb{R}}\mathbf{x} + \mathbf{n}_{\mathbb{R}}. \tag{3.321}$$

For this real-valued model, which is equivalent to the complex-valued model in (3.280), the BLUE, which minimizes the variances of the elements of $\mathbf{x}$ subject to the unbiased constraint is given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}_{\mathbb{R}}^T \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}}^{-1} \mathbf{H}_{\mathbb{R}}\right)^{-1} \mathbf{H}_{\mathbb{R}}^T \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}}^{-1} \mathbf{y}_{\mathbb{R}}. \tag{3.322}$$

Therein, $\mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}}$ is connected with $\underline{\mathbf{C}}_{\mathbf{nn}}$ in the same way as in (2.18). By utilizing this connection, one can easily show that (3.322) corresponds to the estimator in complex notation in Result 3.6. However, for the reasons discussed in [2,39], the complex-valued representation is often favored.

Simplification for Proper Noise

We now analyze Result 3.6 for the case of proper noise. With $\widetilde{\mathbf{C}}_{\mathbf{nn}} = \mathbf{0}^{N_{\mathbf{y}}\times N_{\mathbf{y}}}$ the estimator in (3.307) simplifies to

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{BW}} = &\Big(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H} + \mathbf{H}^T\big(\mathbf{C}_{\mathbf{nn}}^{-1}\big)^*\mathbf{H}^*\Big)^{-1} \\ &\times \Big(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} + \mathbf{H}^T\big(\mathbf{C}_{\mathbf{nn}}^{-1}\big)^*\mathbf{y}^*\Big) \qquad (3.323)\\ =&\big(\mathrm{Re}\big\{\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\big\}\big)^{-1}\mathrm{Re}\big\{\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}\big\}. \qquad (3.324)\end{aligned}$$

Note that this notation is even simpler than the one for improper noise in Result 3.6 and the evaluation of the estimator becomes significantly less complex.

Assuming the special case, where the term $\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}$ is real-valued we obtain from (3.324)

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{BW}} &= \big(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\big)^{-1}\mathrm{Re}\big\{\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}\big\} \qquad (3.325)\\ &= \mathrm{Re}\Big\{\big(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y}\Big\}. \qquad (3.326)\end{aligned}$$

In that case, the BWLUE for real-valued parameter vectors coincides with the real part of the ordinary BLUE in (3.48). Furthermore, it also coincides with the real part of the ordinary BWLUE in (3.63) since the noise is assumed to be proper.

Relation between Number of Measurements and Number of Parameters

Another interesting statement about the estimator can be made concerning the size of the measurement matrix $\mathbf{H}$. Inspecting (3.322) reveals that this estimator is applicable if $\mathbf{H}_{\mathbb{R}} \in \mathbb{R}^{2N_{\mathbf{y}}\times N_{\mathbf{x}}}$ has full column rank and if $2N_{\mathbf{y}} > N_{\mathbf{x}}$. Therefore, only half as many complex-valued measurements are required as there are unknown real-valued parameters. This statement clearly also holds for the BWLUE for real-valued parameter vectors in Result 3.6 since this estimator is equivalent to the one in (3.322).

3.4.2 WLLS Estimator for Real-Valued Parameter Vectors

In the previous section, we showed how the widely linear BWLUE can be modified such that only real-valued estimates are obtained. These ideas are extended to LS estimation in this section. It will turn out that a widely linear version of the LS estimator is obtained naturally. The resulting estimator will be termed the WLLS for real-valued parameter vectors.

The first step for deriving this estimator is to recognize that the LS cost function $J(\mathbf{x})$ in (3.3) is real-valued even for complex $\mathbf{y}$ and $\mathbf{H}$. Hence, it can be written in the form

$$J(\mathbf{x}) = \frac{1}{2}\Big[(\mathbf{y}-\mathbf{H}\mathbf{x})^H(\mathbf{y}-\mathbf{H}\mathbf{x}) + (\mathbf{y}-\mathbf{H}\mathbf{x})^T(\mathbf{y}-\mathbf{H}\mathbf{x})^*\Big]. \tag{3.327}$$

For real-valued $\mathbf{x}$ but complex $\mathbf{H}$ and $\mathbf{y}$, the cost function in (3.327) follows as

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\Big[\mathbf{y}^H\mathbf{y} - \mathbf{y}^H\mathbf{H}\mathbf{x} - \mathbf{x}^T\mathbf{H}^H\mathbf{y} + \mathbf{x}^T\mathbf{H}^H\mathbf{H}\mathbf{x} + \mathbf{y}^T\mathbf{y}^* \\ &\qquad - \mathbf{y}^T\mathbf{H}^*\mathbf{x} - \mathbf{x}^T\mathbf{H}^T\mathbf{y}^* + \mathbf{x}^T\mathbf{H}^T\mathbf{H}^*\mathbf{x}\Big] && (3.328) \\ &= \frac{1}{2}\Big[2\mathbf{y}^H\mathbf{y} - 2\mathbf{y}^H\mathbf{H}\mathbf{x} - 2\mathbf{x}^T\mathbf{H}^H\mathbf{y} + \mathbf{x}^T\big(\mathbf{H}^H\mathbf{H} + \mathbf{H}^T\mathbf{H}^*\big)\mathbf{x}\Big]. && (3.329) \end{aligned}$$

Taking the partial derivative of (3.329) w.r.t. $\mathbf{x}$ yields

$$\begin{aligned} \frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} &= -\mathbf{y}^H\mathbf{H} - \mathbf{y}^T\mathbf{H}^* + \mathbf{x}^T\big(\mathbf{H}^H\mathbf{H} + \mathbf{H}^T\mathbf{H}^*\big) && (3.330) \\ &= -\underline{\mathbf{y}}^H\widetilde{\mathbf{H}} + \mathbf{x}^T\Big(\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\Big). && (3.331) \end{aligned}$$

Note that no Wirtinger calculus for taking the partial derivative is necessary since $\mathbf{x}$ is real-valued. Setting (3.331) equal to zero yields

$$\begin{aligned} \hat{\mathbf{x}}_{\mathrm{LS}} &= \Big(\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\Big)^{-1}\widetilde{\mathbf{H}}^H\underline{\mathbf{y}} && (3.332) \\ &= \mathbf{E}_{\mathrm{LS}}\underline{\mathbf{y}}, && (3.333) \end{aligned}$$

where

$$\mathbf{E}_{\mathrm{LS}} = \Big(\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}\Big)^{-1}\widetilde{\mathbf{H}}^H. \tag{3.334}$$

Due to similar arguments as in (3.310) and (3.312), $\hat{\mathbf{x}}_{\mathrm{LS}}$ fulfills $E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}] = \mathbf{x}$, and the covariance matrix of $\hat{\mathbf{x}}_{\mathrm{LS}}$ simply follows as

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}} &= E_{\mathbf{y}}\Big[(\hat{\mathbf{x}}_{\mathrm{LS}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}])(\hat{\mathbf{x}}_{\mathrm{LS}} - E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}])^H\Big] && (3.335) \\ &= \mathbf{E}_{\mathrm{LS}}\,\underline{\mathbf{C}}_{\mathbf{nn}}\,\mathbf{E}_{\mathrm{LS}}^H. && (3.336) \end{aligned}$$

The derived estimator is summarized in

Result 3.7 (WLLS for Real-Valued Parameter Vectors)

If $\mathbf{x} \in \mathbb{R}^{N_{\mathbf{x}}}$ and $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ are connected via the linear model in (3.280), then the WLLS estimator for real-valued parameter vectors $\hat{\mathbf{x}}_{\mathrm{LS}}$ is given by $\hat{\mathbf{x}}_{\mathrm{LS}} = \mathbf{E}_{\mathrm{LS}}\underline{\mathbf{y}}$, where the estimator matrix $\mathbf{E}_{\mathrm{LS}}$ is defined in (3.334) and $\widetilde{\mathbf{H}}$ is defined in (3.300). This estimator is unbiased in the classical sense, i.e., it fulfills $E_{\mathbf{y}}[\hat{\mathbf{x}}_{\mathrm{LS}}] = \mathbf{x}$, and its covariance matrix is $\mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{LS}}$ is given in (3.336).

We now investigate a special case of this result. Consider the definition of $\widetilde{\mathbf{H}}$ in (3.300), then one can see that (3.333) can be simplified to

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \left(\mathrm{Re}\left\{\mathbf{H}^H\mathbf{H}\right\}\right)^{-1}\left(\mathrm{Re}\left\{\mathbf{H}^H\mathbf{y}\right\}\right). \tag{3.337}$$

Hence, the WLLS estimator for real-valued parameter vectors reduces to the real part of the ordinary LS estimator in (3.5) when the term $\mathbf{H}^H\mathbf{H}$ is real-valued. Furthermore, the BWLUE from Result 3.6 reduces to the WLLS estimator in Result 3.7 by setting the augmented noise covariance matrix equal to the identity matrix.

An extension of Result 3.7 is derived in the following. Replacing the LS cost function in (3.3) by the weighted LS cost function

$$J(\mathbf{x}) = (\mathbf{y} - \mathbf{H}\mathbf{x})^H\mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}) \tag{3.338}$$

allows deriving the weighted widely linear least squares (WWLLS) estimator for real-valued parameter vectors. Assuming that the weighting matrix is a diagonal matrix with positive, real-valued, non-zero and non-infinite diagonal elements makes the derivation a straight forward extension of (3.327)–(3.336) and leads to

Result 3.8 (WWLLS for Real-Valued Parameter Vectors)

If $\mathbf{x} \in \mathbb{R}^{N_\mathbf{x}}$ and $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ are connected via the linear model in (3.280), then the WWLLS estimator for real parameter vectors $\hat{\mathbf{x}}_{\mathrm{WLS}}$ is given by

$$\hat{\mathbf{x}}_{\mathrm{WLS}} = \mathbf{E}_{\mathrm{WLS}}\underline{\mathbf{y}}, \tag{3.339}$$

where the estimator matrix $\mathbf{E}_{\mathrm{WLS}}$ is defined as

$$\mathbf{E}_{\mathrm{WLS}} = \left(\widetilde{\mathbf{H}}^H\underline{\mathbf{W}}\widetilde{\mathbf{H}}\right)^{-1}\widetilde{\mathbf{H}}^H\underline{\mathbf{W}}. \tag{3.340}$$

Here, $\underline{\mathbf{W}}$ is defined as

$$\underline{\mathbf{W}} = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix} \tag{3.341}$$

with $\mathbf{W}$ being a diagonal weighting matrix. This result can further be simplified to

$$\hat{\mathbf{x}}_{\mathrm{WLS}} = \left(\mathrm{Re}\left\{\mathbf{H}^H\mathbf{W}\mathbf{H}\right\}\right)^{-1}\left(\mathrm{Re}\left\{\mathbf{H}^H\mathbf{W}\mathbf{y}\right\}\right). \tag{3.342}$$

The covariance matrix of $\hat{\mathbf{x}}_{\mathrm{WLS}}$ follows as

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{x}}\hat{\mathbf{x}},\mathrm{WLS}} &= E\left[(\hat{\mathbf{x}}_{\mathrm{WLS}} - E[\hat{\mathbf{x}}_{\mathrm{WLS}}])(\hat{\mathbf{x}}_{\mathrm{WLS}} - E[\hat{\mathbf{x}}_{\mathrm{WLS}}])^H\right] && (3.343) \\ &= \mathbf{E}_{\mathrm{WLS}}\,\underline{\mathbf{C}}_{\mathbf{nn}}\,\mathbf{E}_{\mathrm{WLS}}^H. && (3.344) \end{aligned}$$

Note the similarities between the WWLLS estimator in Result 3.8 and the BWLUE for real-valued parameter vectors in Result 3.6.

Example 3.3 (Magnitude Estimation of a Sum of Complex Exponentials)

In this example, real-valued magnitudes of two complex exponentials are estimated from noisy measurements. The measurement at time instance k is written as

$$y[k] = x_1 \exp(j\Omega_1 k) + x_2 \exp(j\Omega_2 k) + n[k], \tag{3.345}$$

where $k = 1, \ldots, N_{\mathbf{y}}$, and x_1 and x_2 are the unknown real-valued magnitudes. Further, Ω_1 and Ω_2 are known normalized angular frequencies. These measurements can be brought into vector/matrix notation as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.346}$$

where $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ is the measurement vector, $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$, and

$$[\mathbf{H}]_{k,l} = \exp(j\Omega_l k), \qquad l = \{1, 2\}. \tag{3.347}$$

The noise $\mathbf{n}$ in (3.346) is chosen to be [32]

$$\mathbf{n} = \sqrt{1 - \rho^2}\mathbf{n}_r + j\rho\mathbf{n}_i, \tag{3.348}$$

where $\mathbf{n}_r$ and $\mathbf{n}_i$ are standard proper Gaussian random vectors of length $N_{\mathbf{y}}$, and $\rho \in [0, 1]$. With this choice, the noise power remains unaffected by the choice of ρ while the improperness of the noise can be adjusted by appropriately choosing ρ. The noise is proper for $\rho = 1/\sqrt{2}$. In the simulations, we choose $\Omega_1 = 0.1$, $\Omega_2 = 0.2$ and $N_{\mathbf{y}} = 20$. The following estimators are considered:

1. The ordinary LS estimator in (3.5), denoted as $\underline{\hat{\mathbf{x}}}_{\text{OLS}}$,
2. an intuitive estimator, which simply takes the real part of the ordinary LS estimator, i.e.,
$$\underline{\hat{\mathbf{x}}} = \text{Re}\{\underline{\hat{\mathbf{x}}}_{\text{OLS}}\}, \tag{3.349}$$
3. the WLLS estimator for real-valued parameter vectors from Result 3.7,
4. the ordinary BWLUE in (3.63), denoted as $\underline{\hat{\mathbf{x}}}_{\text{OBW}}$,
5. the intuitive estimator, which takes the real part of the ordinary BWLUE, i.e.,
$$\underline{\hat{\mathbf{x}}} = \text{Re}\{\underline{\hat{\mathbf{x}}}_{\text{OBW}}\}, \tag{3.350}$$
6. the BWLUE for real-valued parameter vectors from Result 3.6.

The resulting average MSEs (averaged over the elements of $\mathbf{x}$) plotted over ρ are presented in Figure 3.3. The ordinary LS estimator performs worst for all values of ρ (but for $\rho = 1/\sqrt{2}$). Its performance can be increased by considering only the real parts of the estimates. Compared to that, a further increase in performance is achieved by the WLLS estimator from Result 3.7.

The estimators incorporating the improperness of the noise show a performance that strongly depends on ρ. One can see that the BWLUE for real-valued parameter vectors from Result 3.6 outperforms all competing estimators over the whole range of ρ. For $\rho = 1/\sqrt{2}$, which corresponds to proper noise, the BWLUE from Result 3.6 coincides with the WLLS estimator from Result 3.7 since $\mathbf{C_{nn}}$ is a scaled identity matrix and $\tilde{\mathbf{C}}_{\mathbf{nn}} = \mathbf{0}$ in that case.

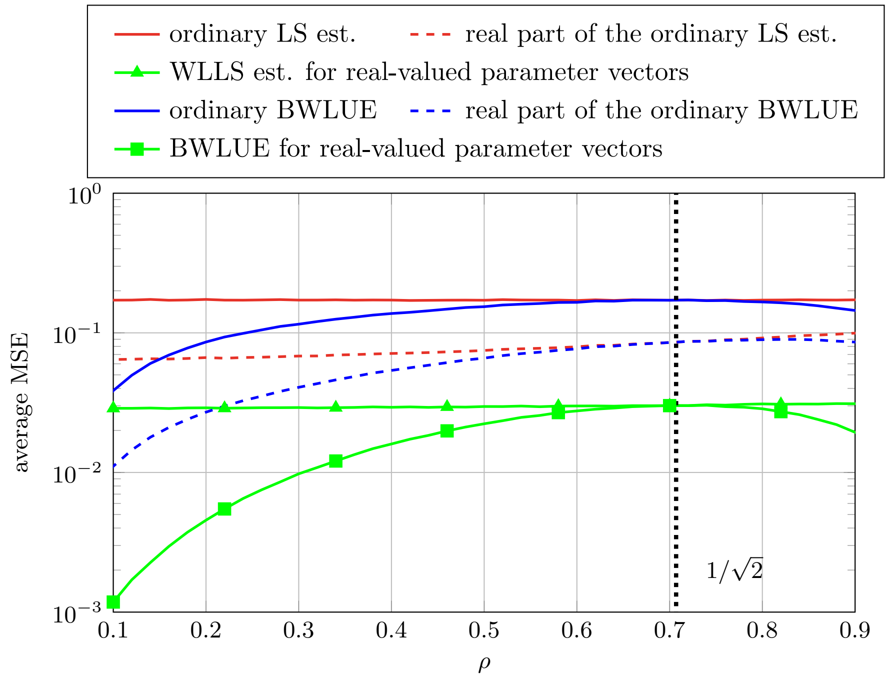

Figure 3.3: Average MSEs of the estimated magnitude values for various estimators. The vertical black line marks the value of $\rho = 1/\sqrt{2}$ where the noise is proper.

The next simulation example deals with the classic non-linear problem of estimating the sampled impulse response of an analog LTI system based on noisy magnitude and phase response measurements. It will turn out that this example impressively shows the performance of the proposed estimators. Furthermore, as part of the example it will be shown how to combine the WLLS estimator in Result 3.7 with the BWLUE for real-valued parameter vectors in Result 3.6. This will reveal a two-step approach that outperforms all the other estimators.

Example 3.4 (Impulse Response Estimation from Noisy Magnitude and Phase Frequency Domain Measurements)

This example deals with the classic non-linear problem of estimating the real-valued impulse response of an LTI system based on separate measurements of the magnitude and phase response [31]. This problem appears in practical scenarios, e.g., when characterizing hydrophones.

Problem Statement

Let the analog real-valued impulse response of an LTI system be denoted as $h(t)$. We are interested in estimating the sampled impulse response $h[n] = h(nT_S)$, where T_S is the sampling time. We assume T_S to be chosen such that the sampling theorem is practically fulfilled, and we furthermore assume the sampled impulse response to be approximately zero after $N_{\mathbf{h}}$ samples. Its samples are put together in the vector $\mathbf{h} \in \mathbb{R}^{N_{\mathbf{h}}}$. The measurements are given by $N_{\mathbf{y}}$ independent magnitude and phase frequency response measurements at equidistant frequencies $f_k = k\Delta f$ with $k = 0, \ldots, N_{\mathbf{y}} - 1$. The true magnitude and phase response values of the analog LTI system at frequency f_k are denoted as A_k and φ_k, respectively, with $A_k \in \mathbb{R}_0^+$ and $\varphi_k \in [0, 2\pi)$, such that the frequency response $H(f_k)$ is given by [31]

$$H(f_k) = A_k e^{j\varphi_k}, \quad k = 0, \ldots, N_{\mathbf{y}} - 1, \tag{3.351}$$

which corresponds to a transformation from polar coordinates to Cartesian coordinates. We now define

$$H_{\mathrm{DC}} = \frac{1}{T_S} H(0), \tag{3.352}$$

$$\mathbf{H}_{\mathrm{AC}} = \frac{1}{T_S} \Big[H(f_1), H(f_2), \ldots, H(f_{N_{\mathbf{y}}-1}) \Big]^T, \tag{3.353}$$

$$\mathbf{H}_{\mathrm{AC,flip}} = \frac{1}{T_S} \Big[H(f_{N_{\mathbf{y}}-1}), H(f_{N_{\mathbf{y}}-2}), \ldots, H(f_1) \Big]^T, \tag{3.354}$$

and

$$\mathbf{H}_{\mathrm{ds}} = \Big[H_{\mathrm{DC}} \quad \mathbf{H}_{\mathrm{AC}}^T \quad \mathbf{H}_{\mathrm{AC,flip}}^H \Big]^T \in \mathbb{C}^{N_D \times 1}, \tag{3.355}$$

with $N_D = 2N_{\mathbf{y}} - 1$. This <u>d</u>ouble-<u>s</u>ided discrete frequency response is connected with the sampled impulse response according to

$$\mathbf{H}_{\mathrm{ds}} = \mathbf{F}_{\mathrm{ds}} \mathbf{h}, \tag{3.356}$$

where $\mathbf{F}_{\mathrm{ds}}$ is the matrix given by the first $N_{\mathbf{h}}$ columns of the DFT matrix of size $N_D \times N_D$. In this example, however, we mainly utilize the <u>s</u>ingle-<u>s</u>ided frequency response $\mathbf{H}_{\mathrm{ss}}$ defined as

$$\mathbf{H}_{\mathrm{ss}} = \Big[H_{\mathrm{DC}} \quad \mathbf{H}_{\mathrm{AC}}^T \Big]^T = \mathbf{F}_{\mathrm{ss}} \mathbf{h}, \tag{3.357}$$

where $\mathbf{F}_{\text{ss}}$ is the $N_{\mathbf{y}} \times N_{\mathbf{h}}$ north-west submatrix of the DFT matrix of size $N_D \times N_D$. While the connection between the sampled impulse response $\mathbf{h}$ and the discrete frequency response in Cartesian coordinates is linear according to (3.356) or (3.357), the relationship between $\mathbf{h}$ and the magnitude- and phase responses A_k and φ_k is non-linear.

Measurement Model

We first concentrate on measurements at frequencies $f_k = k\Delta f$ with $k = 1, \dots, N_{\mathbf{y}} - 1$, and handle the DC measurement later on. The magnitude and phase response measurements at frequency f_k are denoted as $y_k^{(A)}$ and $y_k^{(\varphi)}$, respectively. They are related to A_k and φ_k according to

$$y_k^{(A)} = A_k + n_{A,k} \tag{3.358}$$

and

$$y_k^{(\varphi)} = \varphi_k + n_{\varphi,k} \tag{3.359}$$

for $k = 1, \cdots, N_{\mathbf{y}} - 1$, where $n_{A,k}$ and $n_{\varphi,k}$ denote the corresponding measurement noise variables. In the underlying practical application this work is based on, the measurements as well as the according measurement noise variances were provided by an industry partner. The following assumptions about the noise PDFs are adapted to their method of measuring the magnitude and phase response. $n_{A,k}$ and $n_{\varphi,k}$ are assumed to be statistically independent for all k. Furthermore, we assume $n_{\varphi,k}$ to be zero mean Gaussian with variance $\sigma^2_{\varphi,k}$. Since A_k and $y_k^{(A)}$ have to be positive valued, $n_{A,k}$ cannot be zero mean Gaussian. Thus, in our investigations and simulations we assume $y_k^{(A)}$ for fixed A_k to be Rice distributed according to

$$p(y_k^{(A)}) = \frac{y_k^{(A)}}{\sigma^2_{A,k}} \exp\left(\frac{-\left((y_k^{(A)})^2 + A_k^2\right)}{2\sigma^2_{A,k}} \right) I_0\left(\frac{y_k^{(A)} A_k}{\sigma^2_{A,k}} \right), \tag{3.360}$$

where $I_0(\cdot)$ is the modified Bessel function of first kind with order zero. The resulting mean of $p(y_k^{(A)})$ is denoted as μ_k in the following. For large A_k, the PDF of $n_{A,k}$ is approximately zero mean Gaussian with variance $\sigma^2_{A,k}$, however this is not true for small A_k. Transforming the magnitude and phase response measurements to Cartesian coordinates gives

$$\begin{aligned} y_k &= y_k^{(A)} \mathrm{e}^{j y_k^{(\varphi)}} && (3.361) \\ &= (A_k + n_{A,k}) \mathrm{e}^{j(\varphi_k + n_{\varphi,k})} && (3.362) \\ &= A_k \mathrm{e}^{j\varphi_k} \mathrm{e}^{j n_{\varphi,k}} + n_{A,k} \mathrm{e}^{j\varphi_k} \mathrm{e}^{j n_{\varphi,k}}. && (3.363) \end{aligned}$$

The random variable y_k can also be written as the sum of its mean and a zero mean noise term according to

$$y_k = E[y_k] + n_k. \tag{3.364}$$

From (3.363), the mean $E[y_k]$ becomes

$$E[y_k] = A_k e^{j\varphi_k} E\left[e^{jn_{\varphi,k}}\right] + \mu_{A,k} e^{j\varphi_k} E\left[e^{jn_{\varphi,k}}\right]. \tag{3.365}$$

With $\alpha_k = E\left[e^{jn_{\varphi,k}}\right] = E[\cos(n_{\varphi,k})] = e^{-\sigma^2_{\varphi,k}/2} \in [0,1]$ for $n_{\varphi,k} \sim \mathcal{N}(0,\, \sigma^2_{\varphi,k})$ [40], and the approximation $\mu_{A,k} \approx 0$ (note that $\mu_{A,k}$ depends on the true but unknown magnitude response A_k) we have $E[y_k] \approx \alpha_k H(f_k)$. Therewith, the measurement model (3.364) for $k = 1, \ldots, N_{\mathbf{y}} - 1$ simplifies to

$$y_k \approx \alpha_k H(f_k) + n_k. \tag{3.366}$$

We now analyze the noise term n_k in (3.364). By using the approximation that $n_{A,k}$ has zero mean and variance $\sigma^2_{A,k}$, the variance σ^2_k and pseudo-variance $\widetilde{\sigma}^2_k$ of n_k for $1 \le k \le N_{\mathbf{y}} - 1$ can be approximated by

$$\sigma^2_k = E[(y_k - E[y_k])(y_k - E[y_k])^*] \tag{3.367}$$
$$\approx A^2_k \left(1 - \alpha^2_k\right) + \sigma^2_{A,k} \tag{3.368}$$

and

$$\widetilde{\sigma}^2_k = E[(y_k - E[y_k])(y_k - E[y_k])] \tag{3.369}$$
$$\approx e^{j2\varphi_k} \left(\beta_k A^2_k + \beta_k \sigma^2_{A,k} - A^2_k \alpha^2_k\right), \tag{3.370}$$

respectively. Therein, $\beta_k = E\left[e^{j2n_{\varphi,k}}\right] = E[\cos(2n_{\varphi,k})] = e^{-4\sigma^2_{\varphi,k}/2} \in [0,1]$. The derivations are provided in Appendix E. It is important to note that the noise statistics in (3.368) and (3.370) depend on the true but unknown magnitude and phase response values A_k and φ_k [40,41]. Hence, the true statistics cannot be evaluated without knowledge of the true magnitude and phase response. An obvious option is to replace A_k and φ_k by $y^{(A)}_k$ and $y^{(\varphi)}_k$ in (3.368) and (3.370).

We now turn to the measurement at DC, which can be performed by measuring the steady state system response for a unit step at the input. Instead of a magnitude and a phase the measurement at DC is simply given by a real (positive or negative) scalar value denoted by y_0. We assume the measurement noise at DC to be zero mean Gaussian with variance $\sigma^2_0 = \sigma^2_{A,0}$ and pseudo-variance $\widetilde{\sigma}^2_0 = \sigma^2_0$.

By defining y_{DC}, $\mathbf{y}_{\mathrm{AC}}$, $\mathbf{y}_{\mathrm{AC,flip}}$, $\mathbf{y}_{\mathrm{ds}}$ and $\mathbf{y}_{\mathrm{ss}}$ according to the rules in (3.352)–(3.355) and (3.357), we finally end up with the compact measurement model

$$\mathbf{y}_{\mathrm{ss}} \approx T_S \mathbf{D} \mathbf{F}_{\mathrm{ss}} \mathbf{h} + \mathbf{n}, \tag{3.371}$$

where $\mathbf{D} \in \mathbb{R}^{N_{\mathbf{y}} \times N_{\mathbf{y}}}$ is a diagonal matrix with $[\mathbf{D}]_{1,1} = 1$ and $[\mathbf{D}]_{k+1,k+1} = \alpha_k$ for $k = 1, \ldots, N_{\mathbf{y}} - 1$. Assuming the measurements for different k to be statistically independent, the noise covariance matrix $\mathbf{C}_{\mathbf{nn}}$ and pseudo-covariance matrix $\widetilde{\mathbf{C}}_{\mathbf{nn}}$ follow to be diagonal matrices that can (according to (3.368) and (3.370)) be approximated by $[\mathbf{C}_{\mathbf{nn}}]_{k+1,k+1} = \sigma^2_k$ and $[\widetilde{\mathbf{C}}_{\mathbf{nn}}]_{k+1,k+1} = \widetilde{\sigma}^2_k$ for $k = 0, \ldots, N_{\mathbf{y}} - 1$.

Note that the non-linear connection between the polar measurements and the sampled impulse response has finally been transformed to the model in (3.371) that formally looks like a linear model. Still, it exhibits noise statistics depending on the true magnitude and phase response values A_k and φ_k. The noise statistics consequently depend on the unknown vector $\mathbf{h}$ to be estimated.

Estimators

In contrast to Example 3.3, we now set the number of measurements $N_{\mathbf{y}}$ to be smaller than the number of unknown real-valued parameters $N_{\mathbf{h}}$. While this is not an issue for the BWLUE for real-valued parameter vectors as discussed in Section 3.4.1 (as long as $2N_{\mathbf{y}} \geq N_{\mathbf{h}}$), the ordinary BWLUE fails. Hence, we consider the following estimators [31]:

1. *IDFT based estimator*: The maybe most intuitive and simple estimator is obtained based on (3.356) by replacing $\mathbf{H}_{\mathrm{ds}}$ with the measurements $\mathbf{y}_{\mathrm{ds}}$. An estimate of $\mathbf{h}$ can be obtained by performing an inverse discrete Fourier transform (IDFT) on $\mathbf{y}_{\mathrm{ds}}$ and use the first $N_{\mathbf{h}}$ elements of the result:
$$\hat{\mathbf{h}} = \left(\mathbf{F}^{-1}\widetilde{\mathbf{y}}_{\mathrm{ds}}\right) \odot \mathbf{w}. \tag{3.372}$$
Here, $\mathbf{F}$ is the DFT matrix of size $N_D \times N_D$ and $\mathbf{w} \in \mathbb{R}^{N_D}$ is a windowing vector with ones at the first $N_{\mathbf{h}}$ positions and zeros elsewhere. The operator $\odot$ in (3.372) represents the element-wise multiplication. This estimator is in fact a widely linear estimator since it incorporates the measurements and their complex conjugates in a linear way. It always yields a real-valued $\mathbf{h}$, and since it does not incorporate $\mathbf{D}$ it results in biased estimates.

2. WLLS estimator from Result 3.7: Similar to the IDFT method this estimator does not incorporate the noise statistics. In contrast to the IDFT method, however, the WLLS estimator can also be easily applied if some measurements are missing. This may be helpful in practical applications in which, e.g., it is impossible to measure the frequency response at DC.

3. BWLUE for real-valued parameter vectors from Result 3.6: This estimator is able to incorporate the noise statistics in the form of $\mathbf{C}_{\mathbf{nn}}$ and $\widetilde{\mathbf{C}}_{\mathbf{nn}}$. Since in our application the noise statistics depend on the unknowns A_k and φ_k, we insert the measurements $y_k^{(A)}$ and $y_k^{(\varphi)}$ in (3.368) and (3.370) to obtain approximations of the noise statistics.

4. *Two-step-approach*: Especially when the measurement variances are large, $y_k^{(A)}$ and $y_k^{(\varphi)}$ can deviate heavily from A_k and φ_k, which might lead to bad approximations of the noise statistics in (3.368) and (3.370). We therefore suggest the following two-step estimation approach.

   Step 1: Perform a WLLS estimation, transform the estimated impulse response into frequency domain using a DFT, and use the resulting magnitude and phase response values for approximating the noise statistics in (3.368) and (3.370).

Step 2: Apply the BWLUE for real-valued parameter vectors with the (usually) more precise noise statistics from step 1 to obtain an improved impulse response estimate.

We have to add one comment to the application of the BWLUE for real-valued parameter vectors in this problem: Of course, this estimator requires the augmented noise covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$ to be invertible. Unfortunately, this is not the case due to $\sigma_0^2 = \widetilde{\sigma}_0^2$. However, there exists an easy way to overcome this issue. Consider the real composite model in (3.321) with $\mathbf{H} = T_S\mathbf{D}\mathbf{F}_{\text{ss}}$ (as in (3.371)), and in particular the equation in (3.321) corresponding to the first row of $\text{Im}\{\mathbf{H}\}$. First, due to $\sigma_0^2 = \widetilde{\sigma}_0^2$ the corresponding diagonal entry of $\mathbf{C}_{\mathbf{n}_\mathbb{R}\mathbf{n}_\mathbb{R}}$ is zero (making $\mathbf{C}_{\mathbf{n}_\mathbb{R}\mathbf{n}_\mathbb{R}}$ singular). Second, the first row of $\text{Im}\{\mathbf{H}\}$ is a zero row in our problem, such that the according measurement contains no information about $\mathbf{h}$ at all. Consequently, the corresponding diagonal entry of $\mathbf{C}_{\mathbf{n}_\mathbb{R}\mathbf{n}_\mathbb{R}}$ can be set to any arbitrary non-zero value, which makes the noise covariance matrix $\mathbf{C}_{\mathbf{n}_\mathbb{R}\mathbf{n}_\mathbb{R}}$ and consequently also the augmented noise covariance matrix $\underline{\mathbf{C}}_{\mathbf{nn}}$ invertible.

Simulation Results

Recall that the noise statistics (in Cartesian coordinates) depend on the true magnitude and phase response values. In order to evaluate the MSEs, averaging over the noise statistics is necessary, which implies averaging over the PDF of the impulse response $\mathbf{h}$ in this example. This corresponds to a Bayesian simulation experiment. Hence, the performance measure is the average BMSE (averaged over the elements of the estimated impulse response). For the simulations, the true impulse responses $\mathbf{h}$ with length $N_\mathbf{h} = 12$ are randomly generated by sampling 9 statistically independent samples from a standard proper Gaussian distribution, which are then filtered with a finite impulse response (FIR) filter. Its coefficients are given by

$$\mathbf{c} = \begin{bmatrix} 0.0881 & 0.4408 & 0.4408 & 0.0881 \end{bmatrix}^T. \tag{3.373}$$

This FIR filter corresponds to a low-pass and it takes care that $\mathbf{h}$ shows low-pass characteristics. Next, the DC- and additional 9 noisy magnitude and phase response measurements are generated such that $N_\mathbf{y} = 10$. In the first experiment the noise variance of the phase response measurements is kept constant at $\sigma_{\varphi,k}^2 = 10^{-1}$ for $1 \leq k \leq N_\mathbf{y} - 1$, while the variances $\sigma_{A,k}^2$ are varied between 10^{-5} and 1 for $0 \leq k \leq N_\mathbf{y} - 1$. Since the true impulse responses are generated randomly, the BMSE is used as a performance measure. The resulting average BMSE curves (averaged over the elements of $\mathbf{h}$) plotted over $\sigma_{A,k}^2$ are shown in Figure 3.4. Therein, one can see that the BWLUE for real-valued parameter vectors outperforms the WLLS estimator and the IDFT method significantly. By employing the two-step approach, a further increase in performance is achieved. This two-step approach almost reaches our introduced bound, which is simply generated by applying the BWLUE for real-valued parameter vectors, but with the true A_k and φ_k values inserted in (3.368) and (3.370) to derive the noise statistics.

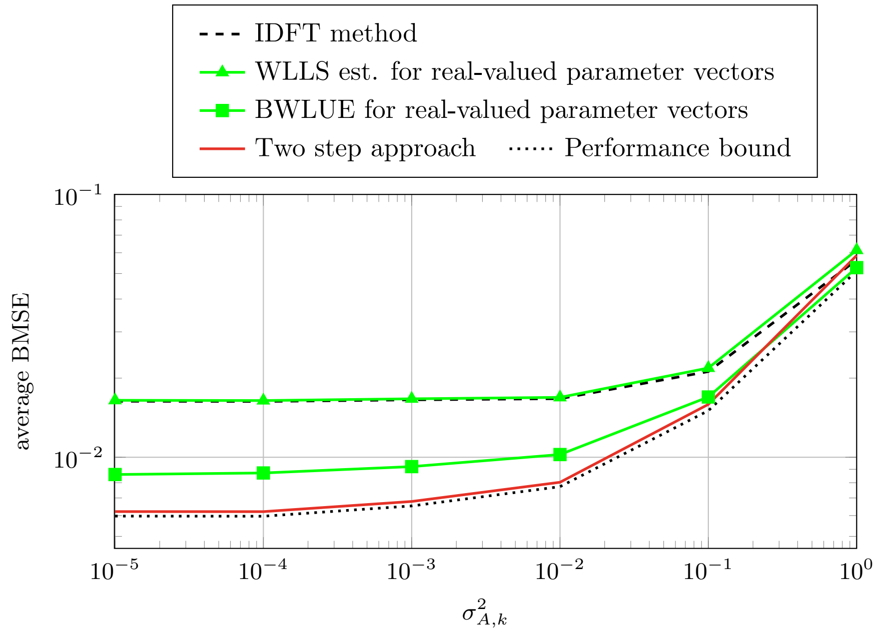

Figure 3.4: Average BMSEs of the estimated impulse response coefficients for various estimators. The noise variances of the phase response measurements are kept constant at $\sigma^2_{\varphi,k} = 10^{-1}$ for $1 \leq k \leq N_{\mathbf{y}} - 1$ while the variances $\sigma^2_{A,k}$ for $0 \leq k \leq N_{\mathbf{y}} - 1$ are varied between 10^{-5} and 1.

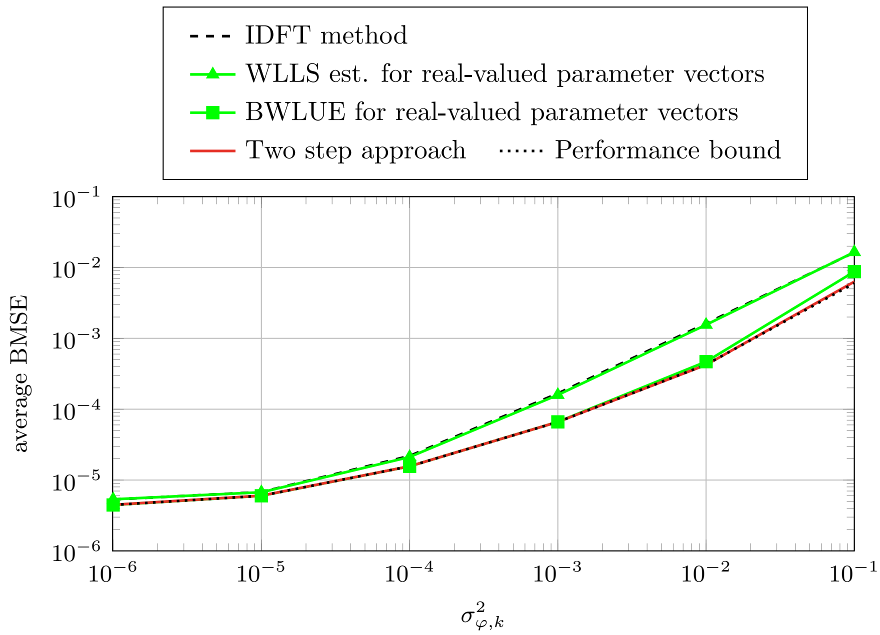

Figure 3.5: Average BMSEs of the estimated impulse response coefficients for various estimators. The variances $\sigma^2_{A,k} = 10^{-4}$ for $0 \leq k \leq N_{\mathbf{y}} - 1$ are kept constant while the variances $\sigma^2_{\varphi,k}$ for $1 \leq k \leq N_{\mathbf{y}} - 1$ are varied between 10^{-6} and 10^{-1}.

In the second experiment $\sigma^2_{A,k} = 10^{-4}$ for $0 \leq k \leq N_{\mathbf{y}} - 1$ is kept constant, while the variances $\sigma^2_{\varphi,k}$ are varied between 10^{-6} and 10^{-1} for $1 \leq k \leq N_{\mathbf{y}} - 1$. The resulting average BMSE curves are shown in Figure 3.5. This figure shows that the BWLUE for real-valued parameter vectors as well as the two-step approach practically reach the bound, except for very large values of $\sigma^2_{\varphi,k}$.

3.5 Parameter Estimation under Model Uncertainties

Consider the linear model in (3.1). There, the measurement matrix $\mathbf{H}$ is assumed to be perfectly known. In practice, this assumption often does not hold. E.g., consider the case where $\mathbf{H}$ is a convolution matrix that is constructed based on an imperfectly measured or estimated signal. The errors in $\mathbf{H}$ are often neglected since they are unknown. However, if statistics of these errors are available, one can improve the estimation performance. The focus of this section lies on accounting for these model uncertainties within the framework of classical estimation.

A practical example where such model errors are present is described in the following. Consider measurements with a sensor, e.g., ultrasound measurements performed with a hydrophone. This hydrophone convolves the ultrasound signals with its impulse response. To reverse the effects of this convolution, the impulse response is measured or estimated. The resulting estimate of the impulse response is affected by an error due to the measurement noise and/or due to the estimation process. Now, the goal is to perform a deconvolution of the measured hydrophone signals in order to obtain the unaltered ultrasound signals. This deconvolution step should account for the fact that the impulse response is affected by an error. A possible way to achieve this is discussed in this section.

State-of-the-Art and Performance Reference

We are considering the model errors as random with known second order statistics but otherwise arbitrary PDF. This is motivated by practical examples such as multiple input multiple output (MIMO) communication channels or beamforming [42–45]. In contrast to the LS estimator and the BLUE, total least squares (TLS) estimation techniques account for these model errors. E.g., for the special case of independent and identically distributed (i.i.d.) Gaussian errors of the elements in $\mathbf{H}$, the maximum likelihood (ML) solution of the TLS problem was analyzed in [46]. However, in many practical applications $\mathbf{H}$ has some sort of structure as it is the case for Toeplitz or Hankel matrices. Then, the model errors are clearly not i.i.d. any more. To deal with these kind of problems, so-called structured total least squares (STLS) techniques have been developed [7–9]. An overview of different TLS and STLS methods can be found in [10–12].

In this section, we compare our novel approach with two iterative algorithms, which serve as performance reference in the simulations later on. The first one, introduced in [46], is an approach for solving the ML problem based on classical expectation-maximization (EM) [47]. This algorithm, referred to as maximum likelihood expectation-maximization (ML-EM) algorithm, treats the model errors as random and allows incorporating the model error variance. By doing so, a uniform variance for every element in $\mathbf{H}$ was assumed. The second one represents an algorithm from the class of STLS approaches and is introduced in [48]. This iterative algorithm is called the structured total least norm (STLN) algorithm. It is capable to deal with structured measurement matrices, and treats the model errors as deterministic but unknown. Hence, it prevents the usage of model error variances.

In the following, an iterative algorithm that is based on the BLUE is discussed. This iterative algorithm allows to combine information about the structure as well as the model error variances. Ultimately, it will turn out that this algorithm can be employed on structured as well as unstructured problems.

Note that a similar iterative application of the BLUE was applied in [49–51] for channel impulse response estimation in wireless communication applications. Compared to these approaches, however, the proposed algorithm is applicable to more general applications with structured or unstructured model uncertainties. In [52] investigations of a similar procedure as the presented algorithm can be found but only for a very simplified model compared to the approach in this section. Because of that, the algorithms presented in [49–52] are not considered as performance reference. Instead, we compare the proposed algorithm with the STLS algorithm in [48], the ML-EM algorithm introduced in [46], and an ideal but only theoretically applicable estimator introduced later on.

3.5.1 System Model

This section describes the underlying model used in the following. In a first step, the elements of the measurement matrix $\mathbf{H}$ are assumed to be unstructured and the model uncertainties therein are assumed to be independent. Afterwards, $\mathbf{H}$ is considered to be a structured convolution matrix built from an estimated or measured impulse response. Hence, $\mathbf{H}$ is a special form of a Toeplitz matrix and, as will be shown, thus allowing correlated model uncertainties.

Unstructured Measurement Matrices

We denote $\hat{\mathbf{H}}$ as the measured or estimated measurement matrix and assume that it comes along with error variances for every entry. The error variances are assembled in a matrix $\mathbf{V} \in \mathbb{R}^{N_\mathbf{y} \times N_\mathbf{x}}$ of the same size as $\hat{\mathbf{H}}$. Furthermore, the errors are assumed to be independent zero mean random variables. The measurements are modelled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = \left(\hat{\mathbf{H}} + \mathbf{B}\right)\mathbf{x} + \mathbf{n}, \tag{3.374}$$

where $\mathbf{H} = \hat{\mathbf{H}} + \mathbf{B}$, with $\hat{\mathbf{H}}$ being the estimated measurement matrix and $\mathbf{B}$ being a zero mean random matrix. The zero mean assumption of the elements in $\mathbf{B}$ implies unbiasedness of the estimates in $\hat{\mathbf{H}}$. In (3.374), $\mathbf{H}$ and $\mathbf{B}$ are unknown, while $\hat{\mathbf{H}}$ is known. We further rewrite (3.374) according to

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{x} + \underbrace{\mathbf{B}\mathbf{x} + \mathbf{n}}_{\mathbf{w}} \tag{3.375}$$

$$= \hat{\mathbf{H}}\mathbf{x} + \mathbf{w}, \tag{3.376}$$

with the new overall noise vector $\mathbf{w} = \mathbf{B}\mathbf{x} + \mathbf{n}$. This noise vector combines the measurement noise with the noise from the model uncertainties. Let $\mathbf{b}_i^T$ be the i^{th} row of $\mathbf{B}$, let w_i be the i^{th} element of $\mathbf{w}$, and let n_i be the i^{th} element of $\mathbf{n}$. Then

$$w_i = \mathbf{b}_i^T \mathbf{x} + n_i. \tag{3.377}$$

Since w_i is evaluated as the scalar product of a vector with zero mean random elements with an unknown but deterministic vector plus n_i, w_i has zero mean and its variance in dependence of the unknown parameter vector $\mathbf{x}$ can be derived as

$$\sigma_{w_i}^2 = [\mathbf{V}]_{i,1} |x_1|^2 + [\mathbf{V}]_{i,2} |x_2|^2 + \cdots + [\mathbf{V}]_{i,N_{\mathbf{x}}} |x_{N_{\mathbf{x}}}|^2 + \sigma_{n_i}^2, \tag{3.378}$$

where x_i is the i^{th} element of $\mathbf{x}$ and where $\sigma_{n_i}^2 = [\mathbf{C}_{\mathbf{nn}}]_{i,i}$ is the variance of n_i. All variances $\sigma_{w_i}^2$ assembled in a covariance matrix yield

$$\mathbf{C}_{\mathbf{ww}} = \operatorname{diag}(\mathbf{V}|\mathbf{x}|^2) + \mathbf{C}_{\mathbf{nn}}, \tag{3.379}$$

where the term $|\mathbf{x}|^2$ represents a column vector of the element-wise absolute squares of the vector $\mathbf{x}$.

Convolution Matrices

We will now assume that $\mathbf{H}$ is a linear convolution matrix constructed from the impulse response $\mathbf{h} \in \mathbb{C}^{N_{\mathbf{h}}}$ of a linear system such that $\mathbf{H}\mathbf{x}$ describes the convolution of the input signal with the impulse response. An extension to other structured measurement matrices is easily possible. Let $\mathbf{H} = \hat{\mathbf{H}} + \mathbf{B}$ have the dimension $N_{\mathbf{y}} \times N_{\mathbf{x}}$, where $N_{\mathbf{y}} = N_{\mathbf{x}} + N_{\mathbf{h}} - 1$. The i^{th} column of the convolution matrices are defined as

$$[\mathbf{H}]_{:,i} = \begin{bmatrix} \mathbf{0}^{(i-1)\times 1} \\ \mathbf{h} \\ \mathbf{0}^{(N_{\mathbf{x}}-i)\times 1} \end{bmatrix}, \quad [\hat{\mathbf{H}}]_{:,i} = \begin{bmatrix} \mathbf{0}^{(i-1)\times 1} \\ \hat{\mathbf{h}} \\ \mathbf{0}^{(N_{\mathbf{x}}-i)\times 1} \end{bmatrix}, \quad [\mathbf{B}]_{:,i} = \begin{bmatrix} \mathbf{0}^{(i-1)\times 1} \\ \mathbf{e} \\ \mathbf{0}^{(N_{\mathbf{x}}-i)\times 1} \end{bmatrix} \tag{3.380}$$

$\forall i = 1, \ldots, N_{\mathbf{x}}$, where $\hat{\mathbf{h}}$ is the estimated impulse response and $\mathbf{e}$ is the unknown error of $\hat{\mathbf{h}}$ with known error covariance matrix $\mathbf{C}_{\mathbf{ee}} \in \mathbb{C}^{N_{\mathbf{h}} \times N_{\mathbf{h}}}$. In this case, the model uncertainties of $\hat{\mathbf{H}}$ are clearly not independent any more, leading to a different calculation of $\mathbf{C}_{\mathbf{ww}}$.

Let $\mathbf{b}_i'$ denote the i^{th} column of $\mathbf{B}$. The subsequent column $\mathbf{b}_{i+1}'$ can be derived by

$$\mathbf{b}_{i+1}' = \underbrace{\begin{bmatrix} \mathbf{0}^{1\times(N_{\mathbf{y}}-1)} & 0 \\ \mathbf{I}^{(N_{\mathbf{y}}-1)\times(N_{\mathbf{y}}-1)} & \mathbf{0}^{(N_{\mathbf{y}}-1)\times 1} \end{bmatrix}}_{\mathbf{D}} \mathbf{b}_i' = \mathbf{D}\mathbf{b}_i', \tag{3.381}$$

i.e., shifting down the elements of $\mathbf{b}_i'$ by one position. With that, the product $\mathbf{Bx}$ in (3.375) follows as

$$\begin{aligned}\mathbf{Bx} &= \mathbf{b}_1' x_1 + \mathbf{b}_2' x_2 + \ldots + \mathbf{b}_{N_\mathbf{x}}' x_{N_\mathbf{x}} && (3.382)\\ &= x_1 \mathbf{b}_1' + x_2 \mathbf{D}\mathbf{b}_1' + \ldots + x_{N_\mathbf{x}} \mathbf{D}^{N_\mathbf{x}-1}\mathbf{b}_1' && (3.383)\\ &= \underbrace{\left(x_1 \mathbf{I} + x_2 \mathbf{D} + \ldots + x_{N_\mathbf{x}} \mathbf{D}^{N_\mathbf{x}-1}\right)}_{\mathbf{P}(\mathbf{x})} \mathbf{b}_1' && (3.384)\\ &= \mathbf{P}(\mathbf{x})\mathbf{b}_1'. && (3.385)\end{aligned}$$

With this result, $\mathbf{w}$ follows as

$$\mathbf{w} = \mathbf{P}(\mathbf{x})\mathbf{b}_1' + \mathbf{n}, \tag{3.386}$$

and its covariance matrix becomes

$$\begin{aligned}\mathbf{C}_{\mathbf{ww}} &= E\left[\left(\mathbf{P}(\mathbf{x})\mathbf{b}_1'\right)\left(\mathbf{P}(\mathbf{x})\mathbf{b}_1'\right)^H\right] + \mathbf{C}_{\mathbf{nn}} && (3.387)\\ &= \mathbf{P}(\mathbf{x})\mathbf{C}_{\mathbf{b}_1'\mathbf{b}_1'}\mathbf{P}(\mathbf{x})^H + \mathbf{C}_{\mathbf{nn}}. && (3.388)\end{aligned}$$

The covariance matrix $\mathbf{C}_{\mathbf{b}_1'\mathbf{b}_1'}$ follows from (3.380) and the covariance matrix of the estimation error $\mathbf{e}$ according to

$$\mathbf{C}_{\mathbf{b}_1'\mathbf{b}_1'} = \begin{bmatrix} \mathbf{C}_{\mathbf{ee}} & \mathbf{0}^{N_\mathbf{h}\times(N_\mathbf{x}-1)} \\ \mathbf{0}^{(N_\mathbf{x}-1)\times N_\mathbf{h}} & \mathbf{0}^{(N_\mathbf{x}-1)\times(N_\mathbf{x}-1)} \end{bmatrix} \in \mathbb{C}^{N_\mathbf{y}\times N_\mathbf{y}}. \tag{3.389}$$

Note that this formulation allows two sources of correlated model errors. The first source of correlation comes from the structure in $\mathbf{H}$. The second source of correlation are correlations within $\mathbf{C}_{\mathbf{ee}}$, which describes the errors in $\hat{\mathbf{h}}$. Hence, the iterative algorithm discussed in the next section is capable of dealing with both kind of correlations.

3.5.2 Iterative Algorithm accounting for Structured and Unstructured Model Errors

An ideal but theoretical estimator for the assumed model is the BLUE applied on the linear model in (3.374). It incorporates the *true* measurement matrix $\mathbf{H}$ and is given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{y}. \tag{3.390}$$

This theoretical estimator is referred to as BLUE *with perfect model knowledge*. Similarly, the BLUE applied on the linear model in (3.376), incorporating the estimated measurement matrix $\hat{\mathbf{H}}$ but the *true* covariance matrix $\mathbf{C}_{\mathbf{ww}}$, follows as

$$\hat{\mathbf{x}} = \left(\hat{\mathbf{H}}^H \mathbf{C}_{\mathbf{ww}}^{-1} \hat{\mathbf{H}}\right)^{-1} \hat{\mathbf{H}}^H \mathbf{C}_{\mathbf{ww}}^{-1} \mathbf{y} \tag{3.391}$$

and it is referred to as BLUE *with perfect knowledge of* $\mathbf{C}_{\mathbf{ww}}$ [50]. The determination of the true $\mathbf{C}_{\mathbf{ww}}$ according to (3.379) or (3.388), however, requires the knowledge of the true parameter vector. To overcome this problem, we propose the iterative algorithm as follows. Its basic idea is to make an initial guess $\hat{\mathbf{x}}^{(0)}$ (the superscript denotes the algorithm's iteration number), which is used to estimate the covariance matrix of $\mathbf{w}$. This estimated covariance matrix is then utilized by an estimator similar to the one in (3.391) to achieve a better estimate of $\mathbf{x}$, which is again used to improve the estimate of the covariance matrix.

The first guess $\hat{\mathbf{x}}^{(0)}$ could for instance origin from an LS estimation which does not incorporate any noise statistics, i.e.,

$$\hat{\mathbf{x}}^{(0)} = \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}}\right)^{-1} \hat{\mathbf{H}}^H \mathbf{y}. \tag{3.392}$$

Then, $\hat{\mathbf{x}}^{(0)}$ is used to estimate $\hat{\mathbf{C}}_{\mathbf{ww}}^{(0)}$ based on (3.379) or (3.388). Hence, we obtain

$$\hat{\mathbf{C}}_{\mathbf{ww}}^{(0)} = \operatorname{diag}(\mathbf{V}|\hat{\mathbf{x}}^{(0)}|^2) + \mathbf{C}_{\mathbf{nn}}, \tag{3.393}$$

for unstructured problems, and

$$\hat{\mathbf{C}}_{\mathbf{ww}}^{(0)} = \mathbf{P}(\hat{\mathbf{x}}^{(0)}) \mathbf{C}_{\mathbf{b}_1' \mathbf{b}_1'} \mathbf{P}(\hat{\mathbf{x}}^{(0)})^H + \mathbf{C}_{\mathbf{nn}} \tag{3.394}$$

for the considered structured problems. This estimated covariance matrix is then incorporated by an estimator similar to (3.391) in order to yield a better estimate $\hat{\mathbf{x}}^{(1)}$, which is then again inserted into (3.393) or (3.394) to obtain $\hat{\mathbf{C}}_{\mathbf{ww}}^{(1)}$, and so on. The generalized update equation is given by

$$\hat{\mathbf{x}}^{(k+1)} = \left(\hat{\mathbf{H}}^H \left(\hat{\mathbf{C}}_{\mathbf{ww}}^{(k)}\right)^{-1} \hat{\mathbf{H}}\right)^{-1} \hat{\mathbf{H}}^H \left(\hat{\mathbf{C}}_{\mathbf{ww}}^{(k)}\right)^{-1} \mathbf{y}. \tag{3.395}$$

Interestingly, the proposed algorithm has similar complexity as the ML-EM and STLN algorithms. It performs a weighting of the measurements according to $\hat{\mathbf{C}}_{\mathbf{ww}}^{(k)}$, which incorporates the model error variances as well as the measurement noise variances. In the case of $\mathbf{H}$ being a convolution matrix, even the covariances of the estimated impulse response are considered in order to improve the estimation. Note that $\hat{\mathbf{C}}_{\mathbf{ww}}^{(k)}$ for both cases is almost always invertible since $\mathbf{C}_{\mathbf{nn}}$ serves as a regularization term in (3.379) and (3.388).

The estimate is unbiased when averaged over the PDF of $\mathbf{n}$ and $\mathbf{B}$, and biased when only averaged over the PDF of $\mathbf{n}$. Let $\mathbf{E}^{(k)} = \left(\hat{\mathbf{H}}^H \left(\hat{\mathbf{C}}_{\mathbf{ww}}^{(k)}\right)^{-1} \hat{\mathbf{H}}\right)^{-1} \hat{\mathbf{H}}^H \left(\hat{\mathbf{C}}_{\mathbf{ww}}^{(k)}\right)^{-1}$ denote the estimator matrix at iteration k. Then, it holds that $\mathbf{E}^{(k)}\hat{\mathbf{H}} = \mathbf{I}$, independent of the estimated parameter vector at the previous iteration cycles. Consequently, the conditional expected vector of $\hat{\mathbf{x}}^{(k+1)}$ for fixed $\mathbf{B}$ follows as

$$\begin{aligned} E_{\mathbf{n}}[\hat{\mathbf{x}}^{(k+1)}|\mathbf{B}] &= E_{\mathbf{n}}[\mathbf{E}^{(k)}\mathbf{y}|\mathbf{B}] && (3.396)\\ &= E_{\mathbf{n}}[\mathbf{E}^{(k)}\hat{\mathbf{H}}\mathbf{x} + \mathbf{E}^{(k)}\mathbf{B}\mathbf{x} + \mathbf{E}^{(k)}\mathbf{n}|\mathbf{B}] && (3.397)\\ &= \mathbf{x} + \mathbf{E}^{(k)}\mathbf{B}\mathbf{x}. && (3.398) \end{aligned}$$

Since $E_{\mathbf{B}}[\mathbf{B}] = \mathbf{0}$, unbiasedness is achieved when averaged over the PDF of $\mathbf{n}$ and $\mathbf{B}$.

Although convergence cannot be ensured, simulations showed that divergence is not an issue for reasonable values of $\mathbf{V}$ and $\mathbf{C_{ee}}$.

A stopping criterion can be implemented in several ways. One possibility is to stop the iterations when $\hat{\mathbf{x}}^{(k)}$ does not significantly change from one iteration to the next. Simulations showed that the major performance gain is usually achieved after the first iteration. Hence, a predefined number of iterations may be utilized instead of a stopping criterion. For this case, the proposed iterative algorithm is summarized in

Result 3.9 (Iterative Algorithm Accounting for Structured and Unstructured Model Errors)

Let $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ and $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$ be connected via (3.374), where $\hat{\mathbf{H}} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a known measured or estimated measurement matrix, $\mathbf{B} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is an unknown zero mean random matrix, and $\mathbf{H} \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}}$ is the unknown, true measurement matrix such that $\mathbf{H} = \hat{\mathbf{H}} + \mathbf{B}$. For unstructured problems, $\mathbf{V} \in \mathbb{R}^{N_\mathbf{y} \times N_\mathbf{x}}$ denotes the matrix containing the variances of the elements of $\mathbf{B}$. For the case of $\mathbf{H}$ being a convolution matrix, $\mathbf{H}$, $\hat{\mathbf{H}}$ and $\mathbf{B}$ are defined in (3.380). Then, the iterative algorithm accounting for model uncertainties is given by

Initialization:
    Initialize $\hat{\mathbf{x}}^{(0)}$ according to (3.392);
    Choose number of iterations N_{iter};
for $k \leftarrow 0$ **to** N_{iter} **do**
    Estimate $\mathbf{C}^{(k)}_{\mathbf{ww}}$ according to (3.393) or (3.394);
    Update $\hat{\mathbf{x}}^{(k)}$ according to (3.395);
end

The resulting estimates at each step k are unbiased when averaged over the PDF of $\mathbf{n}$ and $\mathbf{B}$.

Of course, there exists at least one case where the iterations yield no performance gain. If $\hat{\mathbf{C}}^{(k)}_{\mathbf{ww}}$ is a scaled identity matrix, the proposed algorithm reduces to the ordinary LS estimator, preventing any performance gain. This is, e.g., the case when the following two conditions hold: a) The measurement matrix is unstructured and $\mathbf{V}$ has the same variance at every element. b) the noise covariance matrix $\mathbf{C_{nn}}$ is a scaled identity matrix.

The performance of the algorithm from Result 3.9 compared to the ML-EM algorithm, the STLN algorithm, the BLUE with perfect model knowledge in (3.390) as well as the BLUE with perfect knowledge of $\mathbf{C_{ww}}$ in (3.391), is demonstrated in the next simulation example.

Example 3.5 (Estimation of a Parameter Vector with Model Uncertainties)

In this example, $\mathbf{H} \in \mathbb{R}^{7\times 3}$ is a convolution matrix and describes the discrete convolution of the impulse response $h[n]$ with signal $x[n]$. The vector notations of $h[n]$ and $x[n]$ are given by $\mathbf{h} \in \mathbb{R}^{5\times 1}$ and $\mathbf{x} \in \mathbb{R}^{3\times 1}$, respectively. For the simulations, the impulse response is randomly generated from a Gaussian distribution with mean $E[\mathbf{h}] = \mathbf{0}^{5\times 1}$ and covariance matrix $\mathbf{C_{hh}} = \mathbf{I}^{5\times 5}$. The input signal to be estimated is chosen to be $\mathbf{x} = \begin{bmatrix} 1 & 0.5 & 0.25 \end{bmatrix}^T$. Note that we chose real values for $\mathbf{h}$ and $\mathbf{x}$ since the ML-EM algorithm and the STLN algorithm were designed for the real-valued case. Note, however, that Result 3.9 would also be applicable for complex-valued quantities.

For the first analysis, the noise covariance matrix is a scaled identity matrix $\mathbf{C_{nn}} = \sigma_{\mathbf{n}}^2 \mathbf{I}^{7\times 7}$, where the scaling factor $\sigma_{\mathbf{n}}^2$ is varied between 10^{-8} and 10^{-3}. The impulse response estimation step is assumed to yield zero mean errors with error covariance matrix

$$\mathbf{C_{ee}} = \operatorname{diag}\left(\begin{bmatrix} 10^{-4} & 10^{-5} & 10^{-6} & 10^{-6} & 10^{-6} \end{bmatrix}\right). \tag{3.399}$$

For this model, the proposed algorithm in Result 3.9 is compared with the ideal BLUE in (3.391), the ML-EM algorithm and the STLN algorithm. For the latter one the l_2 norm minimization, a tolerance $\epsilon = 10^{-10}$ and $\mathbf{D} = \mathbf{I}^{5\times 5}$ is chosen. Furthermore, $\mathbf{X}$ (Eq. (2.1) in [48]) is identified to be the first $N_{\mathbf{h}}$ columns of $\mathbf{P}(\mathbf{x})$ in (3.385). For more details on these parameters the reader may refer to [48]. For the ML-EM algorithm $\sigma_{\mathbf{h}}^2$ is set to the mean value of $\mathbf{V}$ [46]. While the STLN algorithm comes with its own termination criterium, for which we choose $\epsilon = 10^{-10}$ [48], the proposed algorithm and the ML-EM algorithm were executed for $N_{\text{iter}} = 10$ iterations. It will turn out later, that N_{iter} could be reduced significantly.

The resulting MSE values averaged over the elements of the MSE vector are presented in Figure 3.6. This figure shows that the proposed algorithm attains the performance given by the BLUE with perfect knowledge of $\mathbf{C_{ww}}$ and outperforms the competing algorithms especially for low $\sigma_{\mathbf{n}}^2$. The performance gain is more than one order of magnitude in MSE for small noise variances. For large noise variances all investigated algorithms perform approximately equal. The reason for this is that the model uncertainties are negligible compared to the large measurement noise samples in that case. For the same reason, the gap between all considered algorithms and the BLUE with perfect model knowledge increases with decreasing noise variance. If one had chosen $\mathbf{C_{ee}}$ to be a scaled identity matrix, the STLN algorithm would have a similar performance as the proposed algorithm for very low noise variances. Large scale numerical simulations showed that the performance gain of the proposed iterative algorithm approximately stays the same for other values of $\mathbf{x}$.

Figure 3.7 shows the convergence behavior of the algorithms for $\sigma_{\mathbf{n}}^2 = 10^{-6}$. First of all, one recognizes that the ML-EM algorithm is not able to significantly improve the estimation accuracy compared to the initial LS estimation in this example. Furthermore, it shows that the STLN algorithm as well as the proposed algorithm achieve most of their performance gains in the first iteration. Hence, this extremely fast convergence allows reducing the number of iterations to one without any significant loss in performance in this example. Further simulations showed that this statement also holds for most of the investigated scenarios.

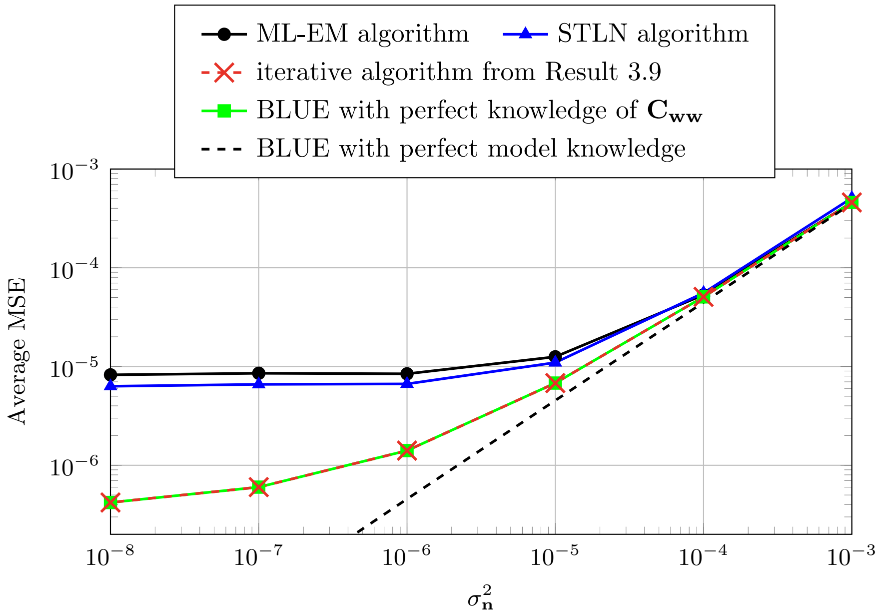

Figure 3.6: Average MSEs of different iterative algorithms, the BLUE with perfect model knowledge in (3.390) as well as the BLUE with perfect knowledge of $\mathbf{C_{ww}}$ in (3.391).

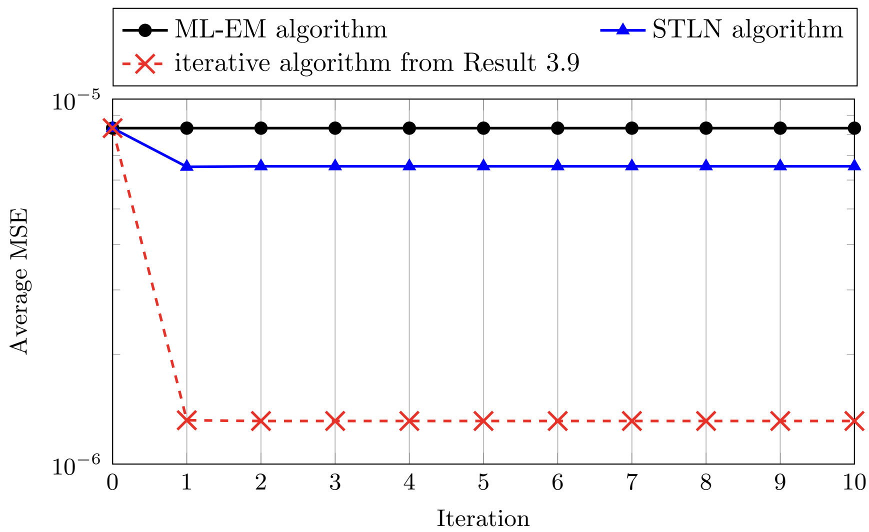

Figure 3.7: Average MSE values plotted over the number of iterations. The noise variance is kept constant at $\sigma_{\mathbf{n}}^2 = 10^{-6}$, and the model error variances are kept constant according to (3.399).

For the next analysis, the noise variance was kept constant at $\sigma_{\mathbf{n}}^2 = 10^{-6}$ and the accuracy of the estimated impulse response was varied by randomly choosing the diagonal elements of $\mathbf{C_{ee}}$ from a uniform distribution between $[0, \zeta]$. The parameter ζ, on the other hand, was varied between $\zeta = 10^{-5}$ and $\zeta = 10^{-2}$. The resulting MSE curves are plotted as a function of ζ in Figure 3.8. Again, the proposed algorithm attains the performance given by the BLUE with perfect knowledge of $\mathbf{C_{ww}}$ and outperforms the ML-EM and STLN algorithms for most values of ζ. For ζ smaller than 10^{-5} all algorithms perform approximately the same. For $\zeta \geq 10^{-2}$ occasional divergence was observed for the proposed algorithm, leading to a decreased MSE performance. Again, the performance gain approximately stays the same for other values of $\mathbf{x}$.

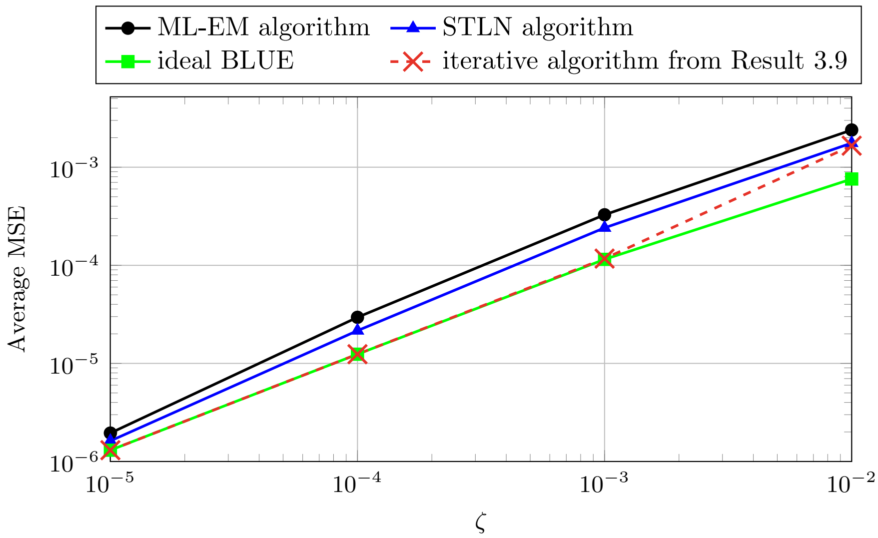

Figure 3.8: Average MSEs of different iterative algorithms and the ideal BLUE in (3.391).

4

Component-Wise Conditionally Unbiased LMMSE and WLMMSE Estimation

This chapter focuses on Bayesian estimation, where the unknown parameter vector is considered to be a random variable whose particular realization has to be estimated [1, 53]. We start by recapitulating well-known Bayesian estimators such as the LMMSE estimator and the WLMMSE estimator. Basically, the main difference between these two estimators is that the LMMSE estimator is linear and only incorporates first and second order statistics, while the WLMMSE estimator is of widely linear form and allows to incorporate augmented first and second order statistics, e.g., the covariance and pseudo-covariance matrices. Both estimators are then analyzed regarding commutation properties and unbiased constraints. These Bayesian estimators utilize a different unbiased constraint than e.g. the classical BLUE. A simulation example will demonstrate the effects of the different unbiased constraints on the estimates. It will be shown that the estimates of the considered Bayesian estimators are conditionally biased.

Based on these findings, a different kind of unbiased constraint is regarded that avoids this conditional bias. These investigations will lead to the CWCU constraints. The CWCU constraints will turn out to be a trade-off between the classical and the usual Bayesian unbiased constraints, which is also demonstrated with a simulation example. In addition, optimal estimators that fulfill the CWCU constraints are derived. These derivations start with the CWCU LMMSE estimator, which is related to the LMMSE estimator. This CWCU LMMSE estimator already derived in [13–15] is extended in this chapter. It is found that the CWCU LMMSE estimator always exists under the usual linear model assumptions, and in the worst case it coincides with the BLUE. However, in a number of practically interesting situations, the CWCU LMMSE estimator is able to outperform the BLUE. We will identify three fundamental scenarios where this is the case.

After that, widely linear CWCU estimators are investigated. These estimators are either novel and unpublished to the best of our knowledge, or published by the author of this work himself. The investigations will lead to the CWCU WLMMSE estimator, which shows strong similarities to the WLMMSE estimator. As for the linear case, the CWCU WLMMSE estimator always exists under the usual linear model assumptions, and in the worst case it coincides with the BWLUE. We will identify several prominent cases for which the CWCU WLMMSE estimator differs from the BWLUE. By doing so, we will strictly distinguish between real and complex-valued parameters since this property significantly influences the expression of the resulting estimators.

Finally, an interesting modification of the CWCU WLMMSE estimator is investigated. This modification, termed the part-wise conditionally unbiased widely linear minimum mean square error (PWCU WLMMSE) estimator, separates real and imaginary parts of the parameter vector and enforces component-wise conditionally unbiasedness on these parts separately. It will turn out that this results in softer constraints of the estimator compared to the CWCU constraints.

Table 4.1 lists the Bayesian estimators that will be discussed and derived in this chapter.

Estimator	***Section***	***Equation/Result***
LMMSE	4.1	(4.14)
WLMMSE	4.1	(4.52)
CWCU LMMSE	4.2	Result 4.1
CWCU WLMMSE for complex-valued parameter vectors	4.3.1	Result 4.2
CWCU WLMMSE for real-valued parameter vectors	4.3.3	Result 4.3
PWCU WLMMSE	4.3.5	

Table 4.1: Bayesian estimators considered in Chapter 4.

In Chapter 3 which discussed classical estimation we utilized the MSE as a performance measure. For Bayesian estimators, however, the BMSE defined as the squared absolute error between the elements of the parameter vector and the estimated parameter vector when averaged over the joint PDF of the measurements and the PDF of the parameter vector, is utilized as performance measure.

4.1 State-of-the-Art

We again consider the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.1}$$

but now $\mathbf{x} \in \mathbb{C}^{N_{\mathbf{x}}}$ is a complex-valued random proper parameter vector, $\mathbf{y} \in \mathbb{C}^{N_{\mathbf{y}}}$ is a complex-valued measurement vector, $\mathbf{H} \in \mathbb{C}^{N_{\mathbf{y}} \times N_{\mathbf{x}}}$ is a complex-valued measurement matrix with full column rank and[4] $N_{\mathbf{x}} < N_{\mathbf{y}}$, and $\mathbf{n} \in \mathbb{C}^{N_{\mathbf{y}}}$ is a complex-valued random proper noise vector with zero mean. We will account for improper $\mathbf{x}$ and $\mathbf{n}$ later.

[4]This requirement is enforced in order to be able to apply the BLUE and in order for the BLUE to differ from the simple matrix inverse. The LMMSE estimator and the WLMMSE estimator do not require this condition.

LMMSE Estimator

We begin with the recapitulation of the derivation of the LMMSE estimator where we focus on the estimation of the i^{th} element of the parameter vector x_i at the beginning. We seek for an affine estimator for x_i of the form

$$\hat{x}_i = \mathbf{e}_i^H \mathbf{y} + b_i, \tag{4.2}$$

where $\mathbf{e}_i \in \mathbb{C}^{N_\mathbf{y}}$ and $b_i \in \mathbb{C}$. The estimator is derived by minimizing the BMSE cost function

$$\begin{aligned}
J(\mathbf{e}_i, b_i) &= E_{\mathbf{y},\mathbf{x}}\Big[|x_i - \hat{x}_i|^2\Big] && (4.3)\\
&= E_{\mathbf{y},\mathbf{x}}\Big[(x_i - \hat{x}_i)(x_i - \hat{x}_i)^H\Big] && (4.4)\\
&= E_{\mathbf{y},\mathbf{x}}\Big[\big(x_i - \mathbf{e}_i^H\mathbf{y} - b_i\big)\big(x_i - \mathbf{e}_i^H\mathbf{y} - b_i\big)^H\Big] && (4.5)\\
&= E_{x_i}[x_i x_i^*] - E_{\mathbf{y},\mathbf{x}}\big[x_i\mathbf{y}^H\big]\mathbf{e}_i - E_{x_i}[x_i]b_i^H - \mathbf{e}_i^H E_{\mathbf{y},\mathbf{x}}\big[\mathbf{y}x_i^H\big]\\
&\quad + \mathbf{e}_i^H E_{\mathbf{y}}\big[\mathbf{y}\mathbf{y}^H\big]\mathbf{e}_i + \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}]b_i^H - b_i E_{x_i}\big[x_i^H\big] + b_i E_{\mathbf{y}}\big[\mathbf{y}^H\big]\mathbf{e}_i + b_i b_i^H. && (4.6)
\end{aligned}$$

Note that the averaging in (4.3) is done w.r.t. the joint PDF of $\mathbf{y}$ and $\mathbf{x}$. Setting the derivative of this cost function w.r.t. b_i equal to zero allows to determine b_i as

$$\begin{aligned}
\frac{\partial J(\mathbf{e}_i, b_i)}{\partial b_i} &= -E_{x_i}\big[x_i^H\big] + E_{\mathbf{y}}\big[\mathbf{y}^H\big]\mathbf{e}_i + b_i^H \stackrel{!}{=} 0 && (4.7)\\
b_i &= E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}]. && (4.8)
\end{aligned}$$

Reinserting this expression into (4.5) yields

$$\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}\Big[\big((x_i - E_{x_i}[x_i]) - \mathbf{e}_i^H(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\big)\big((x_i - E_{x_i}[x_i]) - \mathbf{e}_i^H(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\big)^H\Big] && (4.9)\\
&= \sigma_{x_i}^2 - \mathbf{C}_{x_i\mathbf{y}}\mathbf{e}_i - \mathbf{e}_i^H\mathbf{C}_{\mathbf{y}x_i} + \mathbf{e}_i^H\mathbf{C}_{\mathbf{y}\mathbf{y}}\mathbf{e}_i. && (4.10)
\end{aligned}$$

By setting the derivative of (4.10) w.r.t. $\mathbf{e}_i$ equal to zero, we obtain

$$\begin{aligned}
\frac{\partial J(\mathbf{e}_i)}{\partial \mathbf{e}_i} &= -\mathbf{C}_{x_i\mathbf{y}} + \mathbf{e}_i^H\mathbf{C}_{\mathbf{y}\mathbf{y}} \stackrel{!}{=} \mathbf{0} && (4.11)\\
\mathbf{e}_{\mathrm{L},i}^H &= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}, && (4.12)
\end{aligned}$$

where the index L indicates the LMMSE estimator. We now insert (4.8) and (4.12) into (4.2), yielding

$$\hat{x}_{\mathrm{L},i} = E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{4.13}$$

Note that $\mathbf{C}_{x_i\mathbf{y}}$ corresponds to the i^{th} row of the cross covariance matrix $\mathbf{C}_{\mathbf{x}\mathbf{y}}$. Hence, the LMMSE estimator for the full parameter vector $\mathbf{x}$ immediately follows as

$$\hat{\mathbf{x}}_{\mathrm{L}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{x}\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{4.14}$$

We denote the LMMSE estimation matrix as

$$\mathbf{E}_\mathrm{L} = \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1} \tag{4.15}$$

such that

$$\hat{\mathbf{x}}_\mathrm{L} = E_\mathbf{x}[\mathbf{x}] + \mathbf{E}_\mathrm{L}(\mathbf{y} - E_\mathbf{y}[\mathbf{y}]). \tag{4.16}$$

Since we assume an underlying linear model as in (4.1), the required statistics in (4.14) are given by

$$\mathbf{C}_{\mathbf{xy}} = \mathbf{C}_{\mathbf{xx}}\mathbf{H}^H, \tag{4.17}$$

$$\mathbf{C}_{\mathbf{yy}} = \mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}}, \tag{4.18}$$

$$E_\mathbf{y}[\mathbf{y}] = \mathbf{H}E_\mathbf{x}[\mathbf{x}]. \tag{4.19}$$

Inserting (4.17)–(4.19) into the expression for the LMMSE estimator in (4.14) results in

$$\hat{\mathbf{x}}_\mathrm{L} = E_\mathbf{x}[\mathbf{x}] + \mathbf{C}_{\mathbf{xx}}\mathbf{H}^H\left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}}\right)^{-1}(\mathbf{y} - \mathbf{H}E_\mathbf{x}[\mathbf{x}]) \tag{4.20}$$

$$= E_\mathbf{x}[\mathbf{x}] + \left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H} + \mathbf{C}_{\mathbf{xx}}^{-1}\right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y} - \mathbf{H}E_\mathbf{x}[\mathbf{x}]). \tag{4.21}$$

The equivalence of (4.20) and (4.21) is shown in [1] for the case when $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{nn}}$ are invertible. Since (4.20) does not require invertability of these covariance matrices it can be considered to be a more general expression.

The LMMSE estimator in (4.14) fulfills some optimality criteria listed in the following [1]: The LMMSE estimator is optimal in a BMSE sense

- if the linear model in (4.1) holds and if the prior PDF and the measurement noise PDF are complex proper Gaussian,
- if all terms in (4.1) are real-valued and if $\mathbf{x}$ and $\mathbf{n}$ are Gaussian distributed.

If one of these cases holds, the LMMSE estimator corresponds to the minimum mean square error (MMSE) estimator defined as the mean of the *posterior* PDF $p(\mathbf{x}|\mathbf{y})$

$$\hat{\mathbf{x}} = E_{\mathbf{x}|\mathbf{y}}[\mathbf{x}|\mathbf{y}]. \tag{4.22}$$

Furthermore, the LMMSE estimator also corresponds to the maximum a posteriori (MAP) estimator

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \tag{4.23}$$

in these two cases. This follows from the fact that for Gaussian PDFs, the mode value and the mean value coincide. If none of these two cases hold, the LMMSE estimator is still the best linear (or actually affine) estimator in a BMSE sense. However, non-linear estimators may exist that outperform the LMMSE estimator.

For Bayesian estimators, the covariance matrix of the error $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x}$ is often used as performance measure since it contains the BMSE values of the estimated parameters in its main diagonal. Thus, we will now derive the error covariance matrix for the LMMSE estimator.

The mean of the error $E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}_{\mathrm{L}} - \mathbf{x}]$ is zero since

$$\begin{aligned} E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}_{\mathrm{L}} - \mathbf{x}] &= E_{\mathbf{y},\mathbf{x}}\left[E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - \mathbf{x}\right] && (4.24)\\ &= -E_{\mathbf{x}}[\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}E_{\mathbf{y}}[\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]] && (4.25)\\ &= \mathbf{0}. && (4.26) \end{aligned}$$

With that, the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{L}}$ follows as

$$\begin{aligned} \mathbf{C}_{\mathbf{ee},\mathrm{L}} &= E_{\mathbf{y},\mathbf{x}}\left[\mathbf{e}\mathbf{e}^H\right] && (4.27)\\ &= E_{\mathbf{y},\mathbf{x}}\left[\left(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - \mathbf{x}\right)\left(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - \mathbf{x}\right)^H\right] && (4.28)\\ &= E_{\mathbf{y},\mathbf{x}}\Big[\left(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\right)\\ &\qquad \times \left(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\right)^H\Big] && (4.29)\\ &= \mathbf{C}_{\mathbf{xx}} - \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}} - \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}} + \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}} && (4.30)\\ &= \mathbf{C}_{\mathbf{xx}} - \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}. && (4.31) \end{aligned}$$

The BMSE values of the i^{th} estimate $\hat{x}_{\mathrm{L},i}$ corresponds to the i^{th} diagonal element of the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{L}}$.

As stated before, the MSE performance in general depends on the actual realization of the parameter vector. Consider the i^{th} estimate $\hat{x}_{\mathrm{L},i}$ in (4.13). The conditional MSE under the condition that the parameter vector $\mathbf{x}$ is fixed becomes

$$\mathrm{mse}(\hat{x}_{\mathrm{L},i}|\mathbf{x}) = \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} + \left|x_i - E_{x_i}[x_i] - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\right|^2, \tag{4.32}$$

which is proven in Appendix F. Note that $\mathrm{mse}(\hat{x}_{\mathrm{L},i}|\mathbf{x})$ clearly depends on the actual realization of the parameter vector $\mathbf{x}$. Also note that $\mathrm{mse}(\hat{x}_{\mathrm{L},i}|\mathbf{x})$ consists of two terms. The first term can be shown to be the conditional variance

$$\mathrm{var}(\hat{x}_{\mathrm{L},i}|\mathbf{x}) = \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}. \tag{4.33}$$

The second term turns out to be the absolute square of the conditional bias

$$\begin{aligned} b(\hat{x}_{\mathrm{L},i}|\mathbf{x}) &= E_{\mathbf{y}|\mathbf{x}}[\hat{x}_{\mathrm{L},i} - x_i|\mathbf{x}] && (4.34)\\ &= -x_i + E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]). && (4.35) \end{aligned}$$

Eq. (4.33) and (4.35) are proven in Appendix G. While the conditional variance in (4.33) does not depend on the actual realization of $\mathbf{x}$, the conditional bias in (4.35) clearly does.

The LMMSE estimator commutes over affine transformations, i.e., if the parameter vector is transformed according to $\boldsymbol{\alpha} = \mathbf{B}\mathbf{x} + \mathbf{c}$, then the LMMSE estimator for $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$

is given by $\hat{\boldsymbol{\alpha}}_{\mathrm{L}} = \mathbf{B}\hat{\mathbf{x}}_{\mathrm{L}} + \mathbf{c}$, where $\hat{\mathbf{x}}_{\mathrm{L}}$ is the LMMSE estimator for $\mathbf{x}$ [1], and where $\mathbf{B} \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$. The proof can be found in Appendix H. Note that the dimension $N_{\boldsymbol{\alpha}}$ can be arbitrary. This is an important difference to the investigations in Section 3.1.2.

WLMMSE Estimator

We will now dismiss the assumptions about proper $\mathbf{x}$ and $\mathbf{n}$ and derive a widely linear Bayesian estimator that accounts for improper prior and noise statistics. This will lead us to the WLMMSE estimator for which we utilize the subscript 'WL'.

We seek for a widely linear (actually widely affine) estimator for x_i of the form

$$\begin{aligned} \hat{x}_i &= \mathbf{f}_i^H \mathbf{y} + \mathbf{g}_i^H \mathbf{y}^* + b_i & (4.36) \\ &= \mathbf{e}_i^H \underline{\mathbf{y}} + b_i, & (4.37) \end{aligned}$$

where $\mathbf{e}_i = \begin{bmatrix} \mathbf{f}_i^H & \mathbf{g}_i^H \end{bmatrix}^H \in \mathbb{C}^{2N_{\mathbf{y}}}$ and $b_i \in \mathbb{C}$. The estimator is derived by minimizing the BMSE cost function

$$\begin{aligned} J(\mathbf{e}_i, b_i) &= E_{\mathbf{y},\mathbf{x}}\left[|x_i - \hat{x}_i|^2\right] & (4.38) \\ &= E_{\mathbf{y},\mathbf{x}}\left[(x_i - \hat{x}_i)(x_i - \hat{x}_i)^H\right] & (4.39) \\ &= E_{\mathbf{y},\mathbf{x}}\left[\left(x_i - \mathbf{e}_i^H \underline{\mathbf{y}} - b_i\right)\left(x_i - \mathbf{e}_i^H \underline{\mathbf{y}} - b_i\right)^H\right] & (4.40) \\ &= E_{x_i}\left[x_i x_i^H\right] - E_{\mathbf{y},\mathbf{x}}\left[x_i \underline{\mathbf{y}}^H\right]\mathbf{e}_i - E_{x_i}[x_i] b_i^H - \mathbf{e}_i^H E_{\mathbf{y},\mathbf{x}}\left[\underline{\mathbf{y}} x_i^H\right] \\ &\quad + \mathbf{e}_i^H E_{\mathbf{y}}\left[\underline{\mathbf{y}}\underline{\mathbf{y}}^H\right]\mathbf{e}_i + \mathbf{e}_i^H E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right] b_i^H - b_i E_{x_i}\left[x_i^H\right] + b_i E_{\mathbf{y}}\left[\underline{\mathbf{y}}^H\right]\mathbf{e}_i + b_i b_i^H. & (4.41) \end{aligned}$$

Setting the derivative of this cost function w.r.t. b_i equal to zero allows to identify b_i as

$$\begin{aligned} \frac{\partial J(\mathbf{e}_i, b_i)}{\partial b_i^*} &= -E_{x_i}\left[x_i^H\right] + E_{\mathbf{y}}\left[\underline{\mathbf{y}}^H\right]\mathbf{e}_i + b_i^H \stackrel{!}{=} 0 & (4.42) \\ b_i &= E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]. & (4.43) \end{aligned}$$

Reinserting this expression into (4.40) yields

$$\begin{aligned} J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}\left[\left((x_i - E_{x_i}[x_i]) - \mathbf{e}_i^H\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right)\right)\left((x_i - E_{x_i}[x_i]) - \mathbf{e}_i^H\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right)\right)^H\right] & (4.44) \\ &= \sigma_{x_i}^2 - \mathbf{C}_{x_i\underline{\mathbf{y}}}\mathbf{e}_i - \mathbf{e}_i^H \mathbf{C}_{\underline{\mathbf{y}}x_i} + \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}\mathbf{e}_i, & (4.45) \end{aligned}$$

where $\mathbf{C}_{x_i\underline{\mathbf{y}}} = E_{\mathbf{y},\mathbf{x}}\left[(x_i - E_{x_i}[x_i])\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right)^H\right]$. Setting the derivative of (4.45) w.r.t. $\mathbf{e}_i$ equal to zero leads to

$$\begin{aligned} \frac{\partial J(\mathbf{e}_i)}{\partial \mathbf{e}_i} &= -\mathbf{C}_{x_i\underline{\mathbf{y}}} + \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} = \mathbf{0} & (4.46) \\ \mathbf{e}_{\mathrm{WL},i}^H &= \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}. & (4.47) \end{aligned}$$

The index WL stands for WLMMSE. We are now able to insert (4.43) and (4.47) into (4.37), yielding

$$\hat{x}_{\mathrm{WL},i} = E_{x_i}[x_i] + \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right) \tag{4.48}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right). \tag{4.49}$$

Note that $\mathbf{C}_{x_i\underline{\mathbf{y}}}$ corresponds to the i^{th} row of the augmented cross covariance matrix $\underline{\mathbf{C}}_{\mathbf{xy}}$. Hence, the estimator for the full parameter vector $\mathbf{x}$, which is termed the WLMMSE estimator, can be derived from (4.48) and (4.49) as

$$\hat{\mathbf{x}}_{\mathrm{WL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{C}_{\mathbf{x}\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right) \tag{4.50}$$

$$= E_{\mathbf{x}}[\mathbf{x}] + \begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}} & \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix}\underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right). \tag{4.51}$$

We mainly utilize the WLMMSE estimator for the full augmented parameter vector $\underline{\mathbf{x}}$, which immediately follows as

$$\hat{\underline{\mathbf{x}}}_{\mathrm{WL}} = E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underbrace{\underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}}_{\underline{\mathbf{E}}_{\mathrm{WL}}}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\right), \tag{4.52}$$

where $\underline{\mathbf{E}}_{\mathrm{WL}}$ is the estimator matrix. Since we assume an underlying linear model as in (4.1), the required statistics in (4.52) are given by

$$\underline{\mathbf{C}}_{\mathbf{xy}} = \underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H \tag{4.53}$$

$$\underline{\mathbf{C}}_{\mathbf{yy}} = \underline{\mathbf{H}}\,\underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H + \underline{\mathbf{C}}_{\mathbf{nn}} \tag{4.54}$$

$$E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right] = \underline{\mathbf{H}}E_{\mathbf{x}}[\underline{\mathbf{x}}]. \tag{4.55}$$

Inserting (4.53)–(4.55) into the expression for the WLMMSE estimator in (4.52) results in

$$\hat{\underline{\mathbf{x}}}_{\mathrm{WL}} = E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H\left(\underline{\mathbf{H}}\,\underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H + \underline{\mathbf{C}}_{\mathbf{nn}}\right)^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{H}}E_{\mathbf{x}}[\underline{\mathbf{x}}]\right) \tag{4.56}$$

$$= E_{\mathbf{x}}[\underline{\mathbf{x}}] + \left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}} + \underline{\mathbf{C}}_{\mathbf{xx}}^{-1}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{H}}E_{\mathbf{x}}[\underline{\mathbf{x}}]\right). \tag{4.57}$$

The equivalence of (4.56) and (4.57) can be shown with the help of Woodbury's matrix inversion lemma [54] if $\underline{\mathbf{C}}_{\mathbf{xx}}$ and $\underline{\mathbf{C}}_{\mathbf{nn}}$ are invertible. This is not always the case even if $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{nn}}$ are invertible. To demonstrate such a case we consider a single real-valued parameter with variance $\mathbf{C}_{\mathbf{xx}} = \sigma_x^2$. Since it is real-valued, its pseudo-variance $\widetilde{\sigma}_x^2$ is equal to σ_x^2, resulting in

$$\underline{\mathbf{C}}_{\mathbf{xx}} = \begin{bmatrix}\sigma_x^2 & \widetilde{\sigma}_x^2 \\ \widetilde{\sigma}_x^2 & \sigma_x^2\end{bmatrix} = \begin{bmatrix}\sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2\end{bmatrix}, \tag{4.58}$$

which is clearly not invertible. Hence, the expression of the WLMMSE estimator in (4.56) has to be used in that case.

Note that the WLMMSE estimator in (4.52) reduces to the LMMSE estimator in (4.14) if both the prior PDF as well as the measurement noise PDF are both proper.

There are many similarities between the LMMSE estimator and the WLMMSE estimator. For instance, two of the following three optimality criteria for the WLMMSE estimator are the extensions of the LMMSE's optimality criteria to improper prior knowledge and noise statistics. Assuming $\underline{\mathbf{C}}_{\mathbf{yy}}$ is invertible, then the WLMMSE estimator is optimal in a BMSE sense

- if the linear model in (4.1) holds and if the prior PDF as well as the measurement noise PDF are generalized complex (proper or improper) Gaussian,
- if either $\mathbf{x}$ or $\mathbf{n}$ (or both) in the linear model in (4.1) are real-valued and Gaussian distributed,
- if all terms in (4.1) are real-valued and if $\mathbf{x}$ and $\mathbf{n}$ are Gaussian distributed (then the WLMMSE estimator corresponds to the LMMSE estimator).

In all of these cases, the WLMMSE estimator corresponds to the MMSE estimator in (4.22) and to the MAP estimator in (4.23). Otherwise, the WLMMSE estimator is still the best widely linear (or actually widely affine) estimator in a BMSE sense. However, then non-linear estimators may exist that outperform the WLMMSE estimator.

Statistical measures of the WLMMSE estimator can be derived in a similar manner as it was done for the LMMSE estimator in (4.24)–(4.35). Hence, we will only present the results. The augmented mean $E_{\mathbf{y}}[\hat{\underline{\mathbf{x}}}_{\mathrm{WL}}]$ follows from (4.52) as

$$E_{\mathbf{y}}[\hat{\underline{\mathbf{x}}}_{\mathrm{WL}}] = E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right] - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\big) \tag{4.59}$$

$$= E_{\mathbf{x}}[\underline{\mathbf{x}}]. \tag{4.60}$$

With that, the augmented error covariance matrix $\underline{\mathbf{C}}_{\mathbf{ee},\mathrm{WL}}$ can easily be derived as

$$\underline{\mathbf{C}}_{\mathbf{ee},\mathrm{WL}} = \underline{\mathbf{C}}_{\mathbf{xx}} - \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{yx}}. \tag{4.61}$$

The BMSE values of the i^{th} estimate $\hat{x}_{\mathrm{WL},i}$ corresponds to the i^{th} diagonal element of the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{WL}}$, which is given by the north-west block of $\underline{\mathbf{C}}_{\mathbf{ee},\mathrm{WL}}$. The conditional MSE under the condition that the parameter vector $\mathbf{x}$ is fixed becomes

$$\mathrm{mse}(\hat{x}_{\mathrm{WL},i}|\mathbf{x}) = \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{nn}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{C}_{\underline{\mathbf{y}}x_i} + \left|x_i - E_{x_i}[x_i] - \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{H}}(\underline{\mathbf{x}} - E_{\mathbf{x}}[\underline{\mathbf{x}}])\right|^2. \tag{4.62}$$

The first term in (4.62) can be shown to be the conditional variance

$$\mathrm{var}(\hat{x}_{\mathrm{WL},i}|\mathbf{x}) = \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{nn}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{C}_{\underline{\mathbf{y}}x_i}. \tag{4.63}$$

Further, the second term in (4.62) turns out to be the absolute square of the conditional bias

$$b(\hat{x}_{\mathrm{WL},i}|\mathbf{x}) = E_{\mathbf{y}|\mathbf{x}}[\hat{x}_{\mathrm{WL},i} - x_i|\mathbf{x}] \tag{4.64}$$

$$= -x_i + E_{x_i}[x_i] + \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{H}}(\underline{\mathbf{x}} - E_{\mathbf{x}}[\underline{\mathbf{x}}]). \tag{4.65}$$

Similarly to the LMMSE estimator, the WLMMSE estimator commutes over widely affine transformations, i.e., if the parameter vector is transformed according to $\boldsymbol{\alpha} = \mathbf{B}_1\mathbf{x}+\mathbf{B}_2\mathbf{x}^*+\mathbf{c}$, then the WLMMSE estimator for $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$ is given by $\hat{\boldsymbol{\alpha}}_{\mathrm{WL}} = \mathbf{B}_1\hat{\mathbf{x}}_{\mathrm{WL}}+\mathbf{B}_2\hat{\mathbf{x}}_{\mathrm{WL}}^* + \mathbf{c}$, where $\hat{\mathbf{x}}_{\mathrm{WL}}$ is the WLMMSE estimator for $\mathbf{x}$ [1], and where $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$. The proof can be found in Appendix I. Again, the dimension $N_{\boldsymbol{\alpha}}$ can be arbitrary.

Unbiased Constraints

We now discuss a further important difference between the considered Bayesian and classical estimators concerning their unbiased constraints [1,55–57]. In Section 3.1 we showed that the LS estimator in (3.5), the BLUE in (3.48) as well as the BWLUE in (3.63) fulfill

$$E_{\mathbf{y}}[\hat{\mathbf{x}}] = \mathbf{x} \quad \text{for all possible } \mathbf{x}, \tag{4.66}$$

which is referred to as *classical unbiased constraint.* This result states that for every (deterministic) $\mathbf{x}$, the estimates $\hat{\mathbf{x}}$ are centered around the true parameter vector $\mathbf{x}$. Conversely, the Bayesian LMMSE estimator in (4.14) and the WLMMSE estimators in (4.52) fulfill

$$E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}] = E_{\mathbf{x}}[\mathbf{x}], \tag{4.67}$$

where the integration for $E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}]$ is performed over the joint PDF of $\mathbf{x}$ and $\mathbf{y}$. Eq. (4.67) is referred to as *Bayesian unbiased constraint.* This result states that the considered Bayesian estimators are only "unbiased" when averaged over the PDF of $\mathbf{x}$. Eq. (4.66) can also be formulated in the Bayesian framework. Here, the corresponding problem arises by demanding global conditional unbiasedness, i.e.

$$E_{\mathbf{y}|\mathbf{x}}[\hat{\mathbf{x}}|\mathbf{x}] = \mathbf{x} \quad \text{for all possible } \mathbf{x}. \tag{4.68}$$

The attribute *global* indicates that the condition is made on the whole parameter vector $\mathbf{x}$. This is of importance since another type of conditional unbiased constraint will be discussed later.

Let $\hat{\mathbf{x}} = \mathbf{g}(\mathbf{y})$ be an arbitrary, possible non-linear estimator. For such, the classical unbiased constraint asserts that

$$E_{\mathbf{y}}[\hat{\mathbf{x}}] = \int \mathbf{g}(\mathbf{y})p(\mathbf{y};\mathbf{x})d\mathbf{y} = \mathbf{x} \quad \text{for all possible } \mathbf{x}, \tag{4.69}$$

where $p(\mathbf{y};\mathbf{x})$ is the PDF of vector $\mathbf{y}$ parametrized by the unknown parameter vector $\mathbf{x}$. The Bayesian unbiased constraint on the other hand is

$$E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}} - \mathbf{x}] = \iint (\mathbf{g}(\mathbf{y}) - \mathbf{x})p(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} = \mathbf{0}. \tag{4.70}$$

It will turn out that (4.70) is a much softer requirement than (4.69). However, Bayesian estimators in general allow incorporating prior knowledge about the statistics of $\mathbf{x}$. The global conditional unbiased constraint now reads as

$$E_{\mathbf{y}|\mathbf{x}}[\hat{\mathbf{x}}|\mathbf{x}] = \int \mathbf{g}(\mathbf{y})p(\mathbf{y}|\mathbf{x})d\mathbf{y} = \mathbf{x} \quad \text{for all possible } \mathbf{x}. \tag{4.71}$$

Note that the constricting requirements in (4.68) and (4.71) may prevent the exploitation of prior knowledge about the parameters, and hence lead to a significant reduction in the benefits brought along with the Bayesian framework. Such a case can be demonstrated in the linear model setup by trying to find a linear Bayesian estimator that minimizes the BMSE cost function subject to the constraint (4.68). As shown in Appendix J, the resulting estimator corresponds to the BLUE, which does not utilize any prior knowledge.

To show the effects of the classical unbiased constraint in (4.66) and the Bayesian unbiased constraint in (4.67), an example is regarded now.

Example 4.1 (QPSK Data Estimation (Part 1))

This simple example shall demonstrate the effects of the different unbiased constraints of classical and Bayesian estimators. The task is to estimate zero mean quadrature phase-shift keying (QPSK) data symbols $x_i \in \{\pm 1, \pm j\}$. The measurements are modelled as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{H} \in \mathbb{C}^{10\times 10}$ is given by the first 10 rows and the first 10 columns of a convolution matrix built from the impulse response $\mathbf{h} = \begin{bmatrix} 1.1 & 1 & -0.4 & -0.2 \end{bmatrix}^T$. The noise was chosen to be complex proper Gaussian with covariance matrix $\mathbf{C}_{\mathbf{nn}} = 0.3\,\mathbf{I}^{10\times 10}$. As estimators, the classical BLUE fulfilling (4.66), and the Bayesian LMMSE estimator fulfilling (4.67) are considered. Figure 4.1 shows the relative frequencies of the corresponding estimates in the complex plane, which were generated by performing the estimation task multiple times.

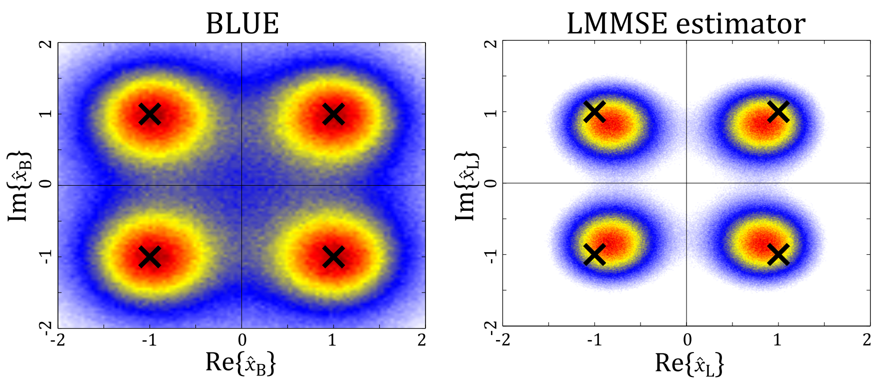

Figure 4.1: Visualization of the relative frequencies of the BLUE and the LMMSE estimator. The black crosses mark the ideal QPSK constellation points.

The estimates of the BLUE are centered around the true constellation points since they are unbiased in the classical sense. In contrast to that, the estimates of the

LMMSE estimator are not centered around the true constellation points. In fact, these estimates are conditionally biased towards the prior mean, which is $\mathbf{0}$. This bias may have to be considered in follow-up processing steps. Note that in Figure 4.1 the BMSE of the LMMSE estimator is clearly below the BMSE of the BLUE.

In the next section, Bayesian estimators are investigated that prevent the shift of the estimates as demonstrated in Figure 4.1.

4.2 Linear CWCU Estimation

In CWCU Bayesian parameter estimation [13–15,58–61], instead of constraining the estimator to be globally unbiased, we aim to achieve 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

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \int g_i(\mathbf{y})p(\mathbf{y}|x_i)d\mathbf{y} = x_i, \tag{4.72}$$

for all possible x_i (and all $i = 1, 2, ..., n$). Note that the CWCU constraints are less stringent than the global conditional unbiased condition in (4.71). Also, it will turn out that a CWCU estimator in many cases allows the incorporation of prior knowledge about the statistical properties of the parameter vector. In the following, we will denote the linear estimator minimizing the BMSE under the CWCU constraints the CWCU LMMSE estimator. The theory of the CWCU LMMSE estimator under linear model assumptions has been discussed in [59–61]. This estimator is of linear (actually affine) form, and it is mainly designed for proper measurement vectors. Its performance and properties will be compared with those of the BLUE and the LMMSE estimator.

It should be noted beforehand, that a CWCU LMMSE estimator cannot outperform the LMMSE estimator in a BMSE sense since it minimizes the BMSE under the additional constraints in (4.72), while the LMMSE estimator's only restriction is the linearity constraint. However, in a number of practically interesting situations, the CWCU LMMSE estimator is able to outperform the BLUE. Furthermore, the CWCU estimators feature their inherent conditional unbiased property that, as it will be shown, preserves the intuitive view of unbiasedness in Bayesian estimation. In order to find linear CWCU estimators that are able to outperform the BLUE, we will investigate certain model assumptions. In particular, we will derive the CWCU LMMSE estimator under the following prerequisites, namely

1. under the assumption of jointly complex Gaussian $\mathbf{x}$ and $\mathbf{y}$,
2. under the linear model assumption with complex Gaussian $\mathbf{x}$ and zero mean noise with known covariance matrix,

3. under the linear model assumption with mutually independent complex (and otherwise arbitrarily distributed) parameters and zero mean noise with known covariance matrix.

We begin with the first case and derive the CWCU LMMSE estimator for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ for which we assign the subscript 'CL'. Note that no assumption about the underlying linear model has to be made in this case.

4.2.1 CWCU LMMSE Estimation under the Jointly Gaussian Assumption

We assume that a vector parameter $\mathbf{x} \in \mathbb{C}^{N_\mathbf{x}}$ is to be estimated based on a measurement vector $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$. As in LMMSE estimation we constrain the estimator to be linear (or actually affine), such that

$$\hat{\mathbf{x}} = \mathbf{E}\mathbf{y} + \mathbf{b}, \qquad \mathbf{E} \in \mathbb{C}^{N_\mathbf{x} \times N_\mathbf{y}}, \mathbf{b} \in \mathbb{C}^{N_\mathbf{x}}. \tag{4.73}$$

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. To show this, let us consider the i^{th} component of the estimator

$$\hat{x}_i = \mathbf{e}_i^H \mathbf{y} + b_i, \tag{4.74}$$

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

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] + b_i. \tag{4.75}$$

A closer inspection of (4.75) 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 function of x_i. For jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ this is the case and we have

$$E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] = E_\mathbf{y}[\mathbf{y}] + \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]), \tag{4.76}$$

where $\mathbf{C}_{\mathbf{y}x_i} = E_{\mathbf{y},\mathbf{x}}[(\mathbf{y} - E_\mathbf{y}[\mathbf{y}])(x_i - E_{x_i}[x_i])^H]$, and $\sigma_{x_i}^2$ is the variance of x_i. Inserting (4.76) into (4.75) produces

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H E_\mathbf{y}[\mathbf{y}] + \mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) + b_i. \tag{4.77}$$

This result reveals that $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled for every x_i if

$$\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1} = 1 \tag{4.78}$$

$$\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma_{x_i}^2 \tag{4.79}$$

and

$$b_i = E_{x_i}[x_i] - \mathbf{e}_i^H E_\mathbf{y}[\mathbf{y}]. \tag{4.80}$$

Now, (4.74), (4.79) and (4.80) allow to simplify the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ as

$$
\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \mathbf{y} + b_i - x_i|^2] && (4.81)\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - (x_i - E_{x_i}[x_i])|^2] && (4.82)\\
&= \mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} \mathbf{e}_i - \underbrace{\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i}}_{\sigma_{x_i}^2} - \underbrace{\mathbf{C}_{x_i\mathbf{y}} \mathbf{e}_i}_{\sigma_{x_i}^2} + \sigma_{x_i}^2 && (4.83)\\
&= \mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2. && (4.84)
\end{aligned}
$$

Finally, the optimization problem is summarized as

$$
\mathbf{e}_{\mathrm{CL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2\right) \quad \text{s.t.} \quad \mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma_{x_i}^2. \tag{4.85}
$$

The optimization problem in (4.85) will now be solved using the Lagrange multiplier method. The Lagrangian cost function is given by

$$
\mathcal{L}(\mathbf{e}_i) = \mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2 + \left(\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} - \sigma_{x_i}^2\right)\lambda + \left(\mathbf{e}_i^T \mathbf{C}_{\mathbf{y}x_i}^* - \sigma_{x_i}^2\right)\lambda^*. \tag{4.86}
$$

Setting the partial derivative of (4.86) w.r.t. $\mathbf{e}_i$ equal to zero allows identifying $\mathbf{e}_i^H$ as

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} &= \mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} + \lambda^* \mathbf{C}_{x_i\mathbf{y}} \overset{!}{=} 0 && (4.87)\\
\mathbf{e}_i^H &= -\lambda^* \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1}. && (4.88)
\end{aligned}
$$

Inserting this result into the constraint in (4.79) allows

$$
\begin{aligned}
-\lambda^* \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i} &= \sigma_{x_i}^2 && (4.89)\\
-\lambda^* &= \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}}. && (4.90)
\end{aligned}
$$

Finally, combining (4.88) and (4.90) produces

$$
\mathbf{e}_{\mathrm{CL},i}^H = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}} \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1}. \tag{4.91}
$$

The full expression for $\hat{x}_{\mathrm{CL},i}$ can be found by combining (4.74), (4.80) and (4.91), which yields

$$
\hat{x}_{\mathrm{CL},i} = E_{x_i}[x_i] + \mathbf{e}_{\mathrm{CL},i}^H(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{4.92}
$$

Using

$$
\mathbf{E}_{\mathrm{CL}} = \begin{bmatrix} \mathbf{e}_{\mathrm{CL},1}^H \\ \mathbf{e}_{\mathrm{CL},2}^H \\ \vdots \\ \mathbf{e}_{\mathrm{CL},N_{\mathbf{x}}}^H \end{bmatrix} \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\mathbf{y}}} \tag{4.93}
$$

immediately leads us to the vector notation of the CWCU LMMSE estimator

$$
\hat{\mathbf{x}}_{\mathrm{CL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CL}}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{4.94}
$$

Note the similarities between the CWCU LMMSE estimator in (4.94) and the LMMSE estimator in (4.14). Also note that according to (4.91), the CWCU LMMSE estimation matrix can be derived as the product of a diagonal matrix times the LMMSE estimation matrix according to

$$\mathbf{E}_{\mathrm{CL}} = \mathbf{D}\mathbf{E}_{\mathrm{L}} \tag{4.95}$$

$$= \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}, \tag{4.96}$$

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

$$[\mathbf{D}]_{i,i} = \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}}. \tag{4.97}$$

It turns out that $[\mathbf{D}]_{i,i}$ is always positive and real-valued. The proof of this statement is straightforward. Consider the definition of $[\mathbf{D}]_{i,i}$ in (4.97). The variance $\sigma_{x_i}^2$ is positive and real-valued per definition. Furthermore, the covariance matrix $\mathbf{C}_{\mathbf{yy}}$ as well as its inverse are Hermitian and positive definite. Multiplying such a matrix with an arbitrary row vector from the left and with the conjugate transpose of the same vector from the right results in a positive and real-valued scalar. Another important fact about $[\mathbf{D}]_{i,i}$ is that

$$[\mathbf{D}]_{i,i} > 1. \tag{4.98}$$

The proof of this statement can be found in Appendix K.

In the following, we derive some performance measures for the CWCU LMMSE estimator. Starting with the mean of the error we have that

$$E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}_{\mathrm{CL}} - \mathbf{x}] = E_{\mathbf{y},\mathbf{x}}\left[E_{\mathbf{x}}[\mathbf{x}] + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - \mathbf{x}\right] \tag{4.99}$$

$$= -E_{\mathbf{x}}[\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]] + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}E_{\mathbf{y}}[\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]] \tag{4.100}$$

$$= \mathbf{0}. \tag{4.101}$$

Therewith, the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{CL}}$ follows as

$$\mathbf{C}_{\mathbf{ee},\mathrm{CL}} = E_{\mathbf{y},\mathbf{x}}\left[\mathbf{e}\mathbf{e}^H\right] \tag{4.102}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - \mathbf{x}\big) \times \big(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - \mathbf{x}\big)^H\Big] \tag{4.103}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\big) \times \big(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\big)^H\Big] \tag{4.104}$$

$$= \mathbf{C}_{\mathbf{xx}} - \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}} - \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}\mathbf{D} + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}\mathbf{D} \tag{4.105}$$

$$= \mathbf{C}_{\mathbf{xx}} - \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}} - \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}\mathbf{D} + \mathbf{D}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}\mathbf{D}. \tag{4.106}$$

By defining a matrix $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}, \tag{4.107}$$

(4.106) simplifies to

$$\mathbf{C}_{\mathbf{ee},\mathrm{CL}} = \mathbf{C}_{\mathbf{xx}} - \mathbf{DM} - \mathbf{MD} + \mathbf{DMD}. \tag{4.108}$$

The BMSE values of the i^{th} estimate $\hat{x}_{\mathrm{CL},i}$ correspond to the i^{th} diagonal element of the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{CL}}$.

4.2.2 CWCU LMMSE Estimation under Linear Model Assumptions

The CWCU LMMSE estimator in (4.94) requires $\mathbf{x}$ and $\mathbf{y}$ to be jointly Gaussian without any further model assumption. We now analyze the situation with an underlying linear model as in (4.1). 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. We could therefore simply insert the adapted covariances

$$\mathbf{C}_{\mathbf{yy}} = \mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}} \tag{4.109}$$

$$\mathbf{C}_{\mathbf{xy}} = \mathbf{C}_{\mathbf{xx}}\mathbf{H}^H \tag{4.110}$$

$$\mathbf{C}_{x_i\mathbf{y}} = \mathbf{C}_{x_i\mathbf{x}}\mathbf{H}^H \tag{4.111}$$

$$\mathbf{C}_{\mathbf{y}x_i} = \mathbf{H}\mathbf{C}_{\mathbf{x}x_i} \tag{4.112}$$

into the CWCU LMMSE estimator. 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, which will be shown in the following. We note that the CWCU LMMSE estimator for the linear model under the assumption of white Gaussian noise has already been derived in [13].

Let $\mathbf{h}_i \in \mathbb{C}^{N_\mathbf{y}}$ be the i^{th} column of $\mathbf{H}$, $\bar{\mathbf{H}}_i \in \mathbb{C}^{N_\mathbf{y} \times N_\mathbf{x}-1}$ the matrix resulting from $\mathbf{H}$ by deleting $\mathbf{h}_i$, and $\bar{\mathbf{x}}_i \in \mathbb{C}^{N_\mathbf{x}-1}$ the vector resulting from $\mathbf{x}$ after deleting x_i. Then the linear model in (4.1) can be rewritten as

$$\mathbf{y} = \mathbf{h}_i x_i + \bar{\mathbf{H}}_i \bar{\mathbf{x}}_i + \mathbf{n}. \tag{4.113}$$

With that, the i^{th} component of $\hat{\mathbf{x}}$ has the form

$$\hat{x}_i = \mathbf{e}_i^H \mathbf{y} + b_i = \mathbf{e}_i^H (\mathbf{h}_i x_i + \bar{\mathbf{H}}_i \bar{\mathbf{x}}_i + \mathbf{n}) + b_i. \tag{4.114}$$

The conditional mean of $\hat{x}_i$ becomes

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H \mathbf{h}_i x_i + \mathbf{e}_i^H \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i] + b_i. \tag{4.115}$$

From (4.115) we can derive conditions that guarantee that the CWCU constraints (4.72) are fulfilled. There are at least the following cases:

1. (4.72) can be fulfilled for all possible x_i if the conditional mean $E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]$ is a linear function of x_i. For complex proper Gaussian $\mathbf{x}$ this condition holds (for all $i = 1, 2, ..., n$).

2. (4.72) can be fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i] = E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]$ for all possible x_i (and all $i = 1, 2, ..., n$), which is true if the elements x_i of $\mathbf{x}$ are mutually independent.

3. (4.72) is fulfilled for all possible x_i (and all $i = 1, 2, ..., n$) if $\mathbf{e}_i^H \mathbf{h}_i = 1$, $\mathbf{e}_i^H \bar{\mathbf{H}}_i = \mathbf{0}^T$, and $b_i = 0$ for $i = 1, 2, \cdots, n$. These constraints and settings correspond to the ones of the BLUE. Consequently, the BLUE fulfills the CWCU constraints.

Solution for Correlated Gaussian Parameters

We now investigate case 1 from above and therefore assume complex proper Gaussian $\mathbf{x}$ with mean $E_{\mathbf{x}}[\mathbf{x}]$ and covariance matrix $\mathbf{C}_{\mathbf{xx}}$. Then we have

$$E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i] = E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]). \tag{4.116}$$

Note that the only requirement on the noise vector so far was its independence on $\mathbf{x}$. Inserting (4.116) into (4.115) produces

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H \mathbf{h}_i x_i + \mathbf{e}_i^H \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{e}_i^H \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) + b_i. \tag{4.117}$$

Again, we obtain two conditions that ensure $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled for every x_i. The first condition is

$$\mathbf{e}_i^H \mathbf{h}_i + \mathbf{e}_i^H \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1} = 1 \tag{4.118}$$

$$\mathbf{e}_i^H \underbrace{\left(\mathbf{h}_i \sigma_{x_i}^2 + \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}\right)}_{\mathbf{C}_{\mathbf{y}x_i}} = \sigma_{x_i}^2. \tag{4.119}$$

The expression in the brackets in (4.119) equals $\mathbf{C}_{\mathbf{y}x_i}$ since

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}x_i} &= E_{\mathbf{y},\mathbf{x}}[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])(x_i - E_{x_i}[x_i])^*] && (4.120)\\
&= E_{\mathbf{y},\mathbf{x}}\left[\left(\mathbf{h}_i(x_i - E_{x_i}[x_i]) + \bar{\mathbf{H}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}\right)(x_i - E_{x_i}[x_i])^*\right] && (4.121)\\
&= \mathbf{h}_i \sigma_{x_i}^2 + \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}. && (4.122)
\end{aligned}$$

Hence, the first condition reads as

$$\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma_{x_i}^2. \tag{4.123}$$

The second condition follows from (4.117) as

$$b_i = \mathbf{e}_i^H \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1} E_{x_i}[x_i] - \mathbf{e}_i^H \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]. \tag{4.124}$$

Incorporating (4.118) into (4.124) yields

$$\begin{aligned}
b_i &= \left(1 - \mathbf{e}_i^H \mathbf{h}_i\right) E_{x_i}[x_i] - \mathbf{e}_i^H \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && (4.125)\\
&= E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}]. && (4.126)
\end{aligned}$$

Together with (4.114), (4.123) and (4.126), the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ can be simplified as

$$\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \mathbf{y} + b_i - x_i|^2] && (4.127)\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H (\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - (x_i - E_{x_i}[x_i])|^2] && (4.128)\\
&= \mathbf{e}_i^H \mathbf{C}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \underbrace{\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i}}_{\sigma_{x_i}^2} - \underbrace{\mathbf{C}_{x_i\mathbf{y}} \mathbf{e}_i}_{\sigma_{x_i}^2} + \sigma_{x_i}^2 && (4.129)\\
&= \mathbf{e}_i^H \mathbf{C}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2. && (4.130)
\end{aligned}$$

Finally, the optimization problem is summarized as

$$\mathbf{e}_{\mathrm{CL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2 \right) \quad \text{s.t.} \quad \mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma_{x_i}^2. \tag{4.131}$$

Most interestingly, this optimization problem equals the one for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ in (4.85). Solving it will formally lead to the same expression for the CWCU LMMSE estimator. However, a significant difference is obtained. By making the assumption about an underlying linear model, the jointly Gaussian assumption of $\mathbf{x}$ and $\mathbf{y}$ can be significantly relaxed. In fact, only the parameter vector $\mathbf{x}$ is required to be Gaussian for resulting in (4.131), while the PDF of the noise $\mathbf{n}$ can be arbitrary. The only requirements on the noise vector are $E_{\mathbf{n}}[\mathbf{n}] = \mathbf{0}$, and $\mathbf{n}$ and $\mathbf{x}$ have to be independent.

Solution for Mutually Independent Parameters

We now investigate case 2 from above and therefore assume that the elements x_i of $\mathbf{x}$ are mutually independent. For mutually independent parameters, it holds that

$$E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i | x_i] = E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]. \tag{4.132}$$

Inserting (4.132) into (4.115) yields

$$E_{\mathbf{y}|x_i}[\hat{x}_i | x_i] = \mathbf{e}_i^H \mathbf{h}_i x_i + \mathbf{e}_i^H \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + b_i. \tag{4.133}$$

The CWCU constraints are fulfilled if

$$\mathbf{e}_i^H \mathbf{h}_i = 1 \tag{4.134}$$

and

$$b_i = -\mathbf{e}_i^H \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i], \tag{4.135}$$

and no further assumptions on the PDF of $\mathbf{x}$ are required [59]. We will now demonstrate that formally the same optimization problem as in (4.85) and (4.131) can be obtained. Adapting (4.122) for mutually independent parameters yields

$$\mathbf{C}_{\mathbf{y}x_i} = \mathbf{h}_i \sigma_{x_i}^2. \tag{4.136}$$

Multiplying (4.134) with $\sigma^2_{x_i}$ and incorporating (4.136) produces

$$\mathbf{e}_i^H \mathbf{C}_{\mathbf{y}x_i} = \sigma^2_{x_i}. \tag{4.137}$$

The second constraint in (4.135) can be rewritten as

$$\begin{aligned} b_i =& -\mathbf{e}_i^H (E_{\mathbf{y}}[\mathbf{y}] - \mathbf{h}_i E_{x_i}[x_i]) && (4.138)\\ =& \, \mathbf{e}_i^H \mathbf{h}_i E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}] && (4.139)\\ =& \, E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\mathbf{y}]. && (4.140) \end{aligned}$$

As in (4.127)–(4.130), one can easily show that the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ for mutually independent parameters yields

$$J(\mathbf{e}_i) = \mathbf{e}_i^H \mathbf{C}_{\mathbf{yy}} \mathbf{e}_i - \sigma^2_{x_i}. \tag{4.141}$$

Hence, the same cost function in (4.141) and the same constraints in (4.137) as for the jointly Gaussian case are obtained. Thus, the CWCU LMMSE estimator for mutually independent parameters formally also equals the CWCU LMMSE estimator for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$.

In [59] we showed that for mutually independent parameters $\mathbf{e}_{\mathrm{CL},i}$ is independent of $\sigma^2_{x_i}$, which can be shown by utilizing (4.91) and (4.136)

$$\begin{aligned} \mathbf{e}_{\mathrm{CL},i}^H =& \frac{\sigma^2_{x_i}}{\sigma^2_{x_i} \mathbf{h}_i^H \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{h}_i \sigma^2_{x_i}} \sigma^2_{x_i} \mathbf{h}_i^H \mathbf{C}_{\mathbf{yy}}^{-1} && (4.142)\\ =& \frac{1}{\mathbf{h}_i^H \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{h}_i} \mathbf{h}_i^H \mathbf{C}_{\mathbf{yy}}^{-1}. && (4.143) \end{aligned}$$

Another fact shown in [59] is that $\mathbf{e}_{\mathrm{CL},i}^H$ for mutually independent parameters can also be brought in the form

$$\mathbf{e}_{\mathrm{CL},i}^H = \frac{1}{\mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i} \mathbf{h}_i^H \mathbf{C}_i^{-1}, \tag{4.144}$$

where

$$\mathbf{C}_i = \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i \bar{\mathbf{x}}_i} \bar{\mathbf{H}}_i^H + \mathbf{C}_{\mathbf{nn}}. \tag{4.145}$$

The proof can be found in Appendix L. Furthermore, we showed that

$$[\mathbf{D}]_{i,i} = (\mathbf{e}_{\mathrm{L},i}^H \mathbf{h}_i)^{-1}, \tag{4.146}$$

where $\mathbf{e}_{\mathrm{L},i}^H$ is the i^{th} row of the LMMSE estimator matrix. This can be easily shown by inserting (4.136) into (4.97), producing

$$\begin{aligned} [\mathbf{D}]_{i,i} =& \frac{\sigma^2_{x_i}}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}} && (4.147)\\ =& \frac{1}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{h}_i} && (4.148)\\ =& \frac{1}{\mathbf{e}_{\mathrm{L},i}^H \mathbf{h}_i} && (4.149)\\ =& (\mathbf{e}_{\mathrm{L},i}^H \mathbf{h}_i)^{-1}. && (4.150) \end{aligned}$$

It therefore holds that $\mathrm{diag}\{\mathbf{E}_{\mathrm{CL}} \mathbf{H}\} = \mathbf{1}$.

Summary

Overall, the investigations of Section 4.2 so far can be summarized in

> **Result 4.1 (CWCU LMMSE Estimator)**
>
> If $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$ are
>
> 1. jointly complex proper Gaussian, or
> 2. connected via the linear model in (4.1) 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
> 3. connected via the linear model in (4.1) and $\mathbf{x}$ has mean $E_{\mathbf{x}}[\mathbf{x}]$, mutually independent elements and covariance matrix $\mathbf{C}_{\mathbf{xx}} = \text{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \cdots, \sigma_{x_n}^2\}$ (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 $i = 1, 2, \cdots, N_{\mathbf{x}}$ is given by (4.94), where the estimator matrix $\mathbf{E}_{\text{CL}}$ is defined in (4.95)–(4.97). The mean of the error $\mathbf{e} = \hat{\mathbf{x}}_{\text{CL}} - \mathbf{x}$ (in the Bayesian sense) is zero, and the error covariance matrix $\mathbf{C}_{\mathbf{ee},\text{CL}}$, which is also the minimum BMSE matrix $\mathbf{M}_{\hat{\mathbf{x}}_{\text{CL}}}$, is provided in (4.108) with $\mathbf{M}$ defined in (4.107). The minimum BMSEs are $\text{Bmse}(\hat{x}_{\text{CL},i}) = [\mathbf{M}_{\hat{\mathbf{x}}_{\text{CL}}}]_{i,i}$.
>
> If none of the three cases is fulfilled, then in the linear model setup a CWCU estimator is available in form of the BLUE, which not necessarily has to correspond to the CWCU LMMSE estimator.

A similar expression for the CWCU LMMSE estimator can be found in [13–15], where the assumption of additive white Gaussian noise (AWGN) has been made.

In the following, some properties of the CWCU LMMSE estimator are detailed.

4.2.3 Discussion of the CWCU LMMSE Estimator

Commonalities between the Three Cases in Result 4.1

We were able to find a CWCU LMMSE estimator deviating from the BLUE for three cases listed in Result 4.1. For the first case it is obvious that

$$E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] = E_{\mathbf{y}}[\mathbf{y}] + \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) \tag{4.151}$$

utilized in (4.76) holds. However, it can be shown that this relation also holds for the other two cases (see Appendix M for proof). Consequently, for all three cases the conditional mean $E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]$ is linear in x_i (which was actually a requirement for finding a linear CWCU estimator).

Similarly, it holds for all three cases that the conditional covariance matrix $\mathbf{C}_{\mathbf{yy}|x_i}$ is given by

$$\mathbf{C}_{\mathbf{yy}|x_i} = \mathbf{C}_{\mathbf{yy}} - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}. \tag{4.152}$$

The proof is presented in Appendix N.

Conditional Properties

In Section 4.1, we demonstrated the dependency of the MSE on $\mathbf{x}$ for the LMMSE and WLMMSE estimators by deriving their conditional MSEs $\mathrm{mse}(\hat{x}_{\mathrm{L},i}|\mathbf{x})$ and $\mathrm{mse}(\hat{x}_{\mathrm{WL},i}|\mathbf{x})$. There, we made the condition on the whole parameter vector $\mathbf{x}$. Since the CWCU constraints contain a condition on x_i only, we analyze several estimators in terms of their conditional mean $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i]$, conditional bias $b(\hat{x}_i|x_i)$, conditional variance $\mathrm{var}(\hat{x}_i|x_i)$ and conditional MSE $\mathrm{mse}(\hat{x}_i|x_i)$.

We begin with the BLUE, which will be analyzed from a Bayesian perspective. This is valid since we showed in Appendix J that the BLUE can also be derived by minimizing the BMSE cost function subject to the global unbiased constraint. Consider the BLUE for x_i in (3.47), then we obtain the following conditional properties:

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{B},i}|x_i] = x_i, \tag{4.153}$$

$$b(\hat{x}_{\mathrm{B},i}|x_i) = 0, \tag{4.154}$$

$$\mathrm{var}(\hat{x}_{\mathrm{B},i}|x_i) = \mathbf{u}_i^H\left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}\mathbf{u}_i, \tag{4.155}$$

$$\mathrm{mse}(\hat{x}_{\mathrm{B},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{B},i}|x_i) = \mathbf{u}_i^H\left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}\mathbf{u}_i. \tag{4.156}$$

The derivation of (4.153)–(4.156) can be found in Appendix O. Note that $\mathrm{var}(\hat{x}_{\mathrm{B},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{B},i})$ and $\mathrm{mse}(\hat{x}_{\mathrm{B},i}|x_i) = \mathrm{mse}(\hat{x}_{\mathrm{B},i})$ holds for the BLUE.

For the derivation of the equivalent properties for the LMMSE estimator we assume that at least one of the three cases mentioned in Result 4.1 holds. Then, the following conditional properties are obtained

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{L},i}|x_i] = [\mathbf{D}]_{i,i}^{-1}x_i + (1 - [\mathbf{D}]_{i,i}^{-1})E_{x_i}[x_i], \tag{4.157}$$

$$b(\hat{x}_{\mathrm{L},i}|x_i) = \Big([\mathbf{D}]_{i,i}^{-1} - 1\Big)(x_i - E_{x_i}[x_i]), \tag{4.158}$$

$$\mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i) = \sigma_{x_i}^2[\mathbf{D}]_{i,i}^{-1}\Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big), \tag{4.159}$$

$$\mathrm{mse}(\hat{x}_{\mathrm{L},i}|x_i) = \sigma_{x_i}^2[\mathbf{D}]_{i,i}^{-1}\Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big) + \Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big)^2|x_i - E_{x_i}[x_i]|^2. \tag{4.160}$$

For the derivations we refer to Appendix P.

These properties of the BLUE and the LMMSE estimator are now compared with those for the CWCU LMMSE estimator (derived in Appendix Q):

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i] = x_i, \tag{4.161}$$

$$b(\hat{x}_{\mathrm{CL},i}|x_i) = 0, \tag{4.162}$$

$$\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i) = \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1), \tag{4.163}$$

$$\mathrm{mse}(\hat{x}_{\mathrm{CL},i}|x_i) = \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). \tag{4.164}$$

Note the interesting connections and similarities between these conditional properties. First of all, the conditional mean and the conditional bias of the CWCU LMMSE estimator of course correspond to those of the BLUE. The LMMSE estimator on the other hand is conditionally biased as it can be seen in (4.158). This bias origins from (4.157), which reveals that the LMMSE estimates are shifted towards the prior mean $E_{x_i}[x_i]$ since $[\mathbf{D}]_{i,i} > 1$. For the special case of a zero mean parameter $E_{x_i}[x_i] = 0$, it holds that

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i] = [\mathbf{D}]_{i,i} E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{L},i}|x_i]. \tag{4.165}$$

Considering the conditional variance, we observe that $\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i)$ is closely related to $\mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i)$ according to

$$\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i) = [\mathbf{D}]_{i,i}^2 \mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i). \tag{4.166}$$

Consequently, it holds that $\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i) > \mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i)$ [59]. A similar relation between the conditional MSEs of the CWCU LMMSE estimator and the LMMSE estimator cannot be found since, in contrast to $\mathrm{mse}(\hat{x}_{\mathrm{CL},i}|x_i)$, $\mathrm{mse}(\hat{x}_{\mathrm{L},i}|x_i)$ clearly depends on the actual realization of x_i. However, for the BMSEs again a simple relation can be found [59]:

$$
\begin{aligned}
\mathrm{Bmse}(\hat{x}_{\mathrm{L},i}) &= E_{x_i}[\mathrm{mse}(\hat{x}_{\mathrm{L},i}|x_i)] && (4.167)\\
&= \sigma_{x_i}^2 [\mathbf{D}]_{i,i}^{-1}\left(1 - [\mathbf{D}]_{i,i}^{-1}\right) + \left(1 - [\mathbf{D}]_{i,i}^{-1}\right)^2 \sigma_{x_i}^2 && (4.168)\\
&= \sigma_{x_i}^2 \left(1 - [\mathbf{D}]_{i,i}^{-1}\right)\left([\mathbf{D}]_{i,i}^{-1} + 1 - [\mathbf{D}]_{i,i}^{-1}\right) && (4.169)\\
&= \sigma_{x_i}^2 \left(1 - [\mathbf{D}]_{i,i}^{-1}\right). && (4.170)
\end{aligned}
$$

For the CWCU LMMSE estimator we trivially obtain

$$\mathrm{Bmse}(\hat{x}_{\mathrm{CL},i}) = \mathrm{mse}(\hat{x}_{\mathrm{CL},i}|x_i) = \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). \tag{4.171}$$

Comparing (4.170) with (4.171) reveals the following relation

$$\mathrm{Bmse}(\hat{x}_{\mathrm{CL},i}) = [\mathbf{D}]_{i,i} \mathrm{Bmse}(\hat{x}_{\mathrm{L},i}), \tag{4.172}$$

consequently $\mathrm{Bmse}(\hat{x}_{\mathrm{CL},i}) > \mathrm{Bmse}(\hat{x}_{\mathrm{L},i})$ holds. Hence, the loss in BMSE performance compared to the LMMSE estimator directly follows from the diagonal matrix $\mathbf{D}$. We emphasize again that for the derivation of the previous results, we assumed that at least one of the three cases mentioned in Result 4.1 holds.

Similar investigations for the BLUE directly lead to $\mathrm{mse}(\hat{x}_{\mathrm{B},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{B},i}|x_i) = \mathrm{Bmse}(\hat{x}_{\mathrm{B},i})$.

Relation to the BLUE and to the LMMSE Estimator

In the previous section it turned out that the CWCU LMMSE estimator is closely related to the LMMSE estimator. Actually, this relation has already been indicated in Section 4.2 (particularly in (4.95)), where it was shown that the CWCU LMMSE estimator matrix can be derived by multiplying the LMMSE estimator matrix with a real-valued diagonal matrix. For the special case of zero mean parameters $E_{\mathbf{x}}[\mathbf{x}] = \mathbf{0}$, the CWCU LMMSE estimates are scaled LMMSE estimates as

$$\hat{\mathbf{x}}_{\mathrm{CL}} = \mathbf{E}_{\mathrm{CL}}\mathbf{y} = \mathbf{D}\mathbf{E}_{\mathrm{L}}\mathbf{y} = \mathbf{D}\hat{\mathbf{x}}_{\mathrm{L}}. \tag{4.173}$$

For the case where the linear model in (4.1) holds and where the parameters are mutually independent, the diagonal elements of $\mathbf{D}$ directly follow from the LMMSE estimator matrix according to (4.146).

The relation between the CWCU LMMSE estimator and the BLUE is not that obvious. However, in some special cases, these two estimators are actually equivalent. For example, this is the case when there is only one parameter to be estimated, i.e., $N_{\mathbf{x}} = 1$. Then, the CWCU constraints correspond to the global conditional unbiasedness $E_{\mathbf{y}|\mathbf{x}}[\hat{\mathbf{x}}|\mathbf{x}] = \mathbf{x}$ in (4.68), which is also fulfilled by the BLUE. Another case is when $\mathbf{C}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{nn}}$ and $\mathbf{H}$ are all diagonal matrices, which is proven in Appendix R. And as already said in Result 4.1, if non of the three mentioned cases is fulfilled, then in the linear model setup a CWCU estimator is available in form of the BLUE, which not necessarily has to correspond to the CWCU LMMSE estimator.

Transformation Analysis

The CWCU LMMSE estimator will in general not commute over affine transformations of the form [59]

$$\boldsymbol{\alpha} = \mathbf{B}\mathbf{x} + \mathbf{c}. \tag{4.174}$$

This is shown in the following way. Let the i^{th} row of the transformation matrix $\mathbf{B}$ be denoted by $\mathbf{b}_i^H$ such that

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^H \\ \mathbf{b}_2^H \\ \vdots \\ \mathbf{b}_{N_{\boldsymbol{\alpha}}}^H \end{bmatrix} \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}, \tag{4.175}$$

and let α_i and c_i be the i^{th} elements of $\boldsymbol{\alpha}$ and $\mathbf{c}$, respectively. We now assume that one of the three cases in Result 4.1 holds for $\boldsymbol{\alpha}$ instead of $\mathbf{x}$ such that the CWCU LMMSE estimator for α_i follows as

$$\hat{\alpha}_{\mathrm{CL},i} = E_{\alpha_i}[\alpha_i] + \frac{\sigma_{\alpha_i}^2}{\mathbf{C}_{\alpha_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}\alpha_i}}\mathbf{C}_{\alpha_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{4.176}$$

From (4.174), the following relations can be derived

$$\alpha_i = \mathbf{b}_i^H \mathbf{x} + c_i, \tag{4.177}$$

$$E_{\alpha_i}[\alpha_i] = \mathbf{b}_i^H E_{\mathbf{x}}[\mathbf{x}] + c_i, \tag{4.178}$$

$$\sigma_{\alpha_i}^2 = \mathbf{b}_i^H \mathbf{C}_{\mathbf{xx}} \mathbf{b}_i, \tag{4.179}$$

$$\mathbf{C}_{\mathbf{y}\alpha_i} = E_{\mathbf{y},\alpha_i}[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])(\alpha_i - E_{\alpha_i}[\alpha_i])^*], \tag{4.180}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\big(\mathbf{b}_i^H(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big)^*\Big], \tag{4.181}$$

$$= \mathbf{C}_{\mathbf{yx}} \mathbf{b}_i. \tag{4.182}$$

Inserting (4.178)–(4.182) into the CWCU LMMSE estimator in (4.176) produces

$$\hat{\alpha}_{\mathrm{CL},i} = \mathbf{b}_i^H E_{\mathbf{x}}[\mathbf{x}] + c_i + \frac{\mathbf{b}_i^H \mathbf{C}_{\mathbf{xx}} \mathbf{b}_i}{\mathbf{b}_i^H \mathbf{C}_{\mathbf{xy}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{yx}} \mathbf{b}_i} \mathbf{b}_i^H \mathbf{C}_{\mathbf{xy}} \mathbf{C}_{\mathbf{yy}}^{-1} (\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \tag{4.183}$$

$$= \mathbf{b}_i^H \left(E_{\mathbf{x}}[\mathbf{x}] + \frac{\mathbf{b}_i^H \mathbf{C}_{\mathbf{xx}} \mathbf{b}_i}{\mathbf{b}_i^H \mathbf{C}_{\mathbf{xy}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{yx}} \mathbf{b}_i} \mathbf{C}_{\mathbf{xy}} \mathbf{C}_{\mathbf{yy}}^{-1} (\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \right) + c_i. \tag{4.184}$$

As it can be easily verified, the expression in the brackets in (4.184) does not correspond to the CWCU LMMSE estimator for $\mathbf{x}$. Hence, the CWCU LMMSE estimator in general does not commute over affine transformations.

However, there exists at least one exception. Consider the case of diagonal transformation matrices

$$\mathbf{B} = \mathrm{diag}\Big\{\Big[b_1 \quad b_2 \quad \ldots \quad b_{N_{\mathbf{x}}}\Big]\Big\} \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\mathbf{x}}} \tag{4.185}$$

with non-zero diagonal elements. This requirement ensures invertability of $\mathbf{B}$. By constraining the transformation matrix to be diagonal, the expressions in (4.177)–(4.182) read as

$$\alpha_i = b_i x_i + c_i \tag{4.186}$$

$$E_{\alpha_i}[\alpha_i] = b_i E_{x_i}[x_i] + c_i \tag{4.187}$$

$$\sigma_{\alpha_i}^2 = |b_i|^2 \sigma_{x_i}^2 \tag{4.188}$$

$$\mathbf{C}_{\mathbf{y}\alpha_i} = E_{\mathbf{y},\alpha_i}[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])(\alpha_i - E_{\alpha_i}[\alpha_i])^*] \tag{4.189}$$

$$= E_{\mathbf{y},x_i}[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])(b_i(x_i - E_{x_i}[x_i]))^*] \tag{4.190}$$

$$= \mathbf{C}_{\mathbf{y}x_i} b_i^*. \tag{4.191}$$

Now, inserting (4.187)–(4.191) into the CWCU LMMSE estimator in (4.176) and utilizing the fact that the diagonal elements are non-zero results in

$$\hat{\alpha}_{\mathrm{CL},i} = b_i E_{x_i}[x_i] + c_i + \frac{|b_i|^2 \sigma_{x_i}^2}{b_i \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i} b_i^*} b_i \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} (\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \tag{4.192}$$

$$= b_i \left(E_{x_i}[x_i] + \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{C}_{\mathbf{y}x_i}} \mathbf{C}_{x_i\mathbf{y}} \mathbf{C}_{\mathbf{yy}}^{-1} (\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \right) + c_i \tag{4.193}$$

$$= b_i \hat{x}_{\mathrm{CL},i} + c_i. \tag{4.194}$$

In conclusion, the CWCU LMMSE estimator commutes over affine transformations with diagonal and invertible transformation matrices.

After these theoretical investigations of the CWCU LMMSE estimator, we demonstrate some practical examples in the following. We begin with an extension of Example 4.1.

Example 4.2 (QPSK Data Estimation (Part 2))

For the exact same setup as in Example 4.1, the CWCU LMMSE estimator is compared with the BLUE and the LMMSE estimator. Note that the linear model in (4.1) holds and the elements of the parameter vector are statistically independent. Figure 4.2 shows the relative frequencies of the corresponding estimates in the complex plane.

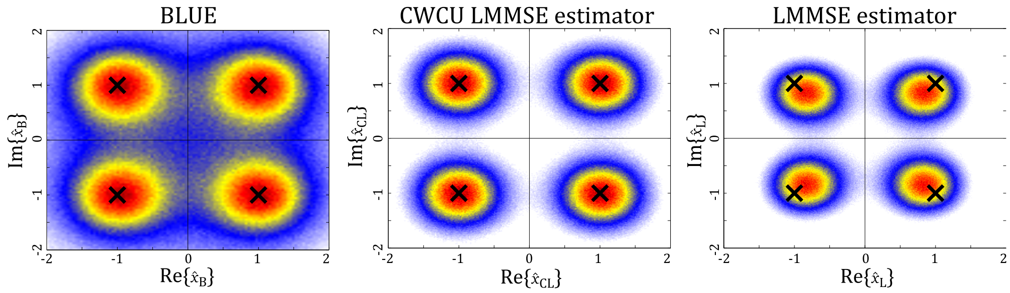

Figure 4.2: Visualization of the relative frequencies of the BLUE, the CWCU LMMSE estimator and the LMMSE estimator. The black crosses mark the ideal QPSK constellation points.

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 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 for each element of the parameter vector. The CWCU constraints prevent this bias introduced by the LMMSE estimator, while prior knowledge about the data can still be incorporated. This prior knowledge effectively reduces the BMSE compared to the BLUE. Hence, Figure 4.2 nicely demonstrates the effects of the CWCU constraints as a trade-off between classical and Bayesian LMMSE estimation. Although the BMSEs of the CWCU LMMSE estimator and the LMMSE estimator differ, it will turn out in the next example that their corresponding log-likelihood ratio (LLR) values and consequently the bit error ratio (BER) coincide [58].

Example 4.3 (Log-Likelihood Ratio Evaluation of a CWCU LMMSE QAM Data Estimator)

In wireless communications, channel coding is applied that introduces redundancy in order to improve the BER behaviour [62]. Consequently, a decoder at the receiver side is required for decoding the bit stream. In soft decoding, the decoder requires the LLR of every bit based on the estimated data symbols [63]. In Example 4.2 we showed that these estimated data symbols strongly depend on whether using an LMMSE estimator or a CWCU LMMSE estimator. However, it will turn out that both estimators yield identical LLRs, resulting also in identical BER performance [58].

After estimating the data symbols as it was done exemplarily in Example 4.2, the LLR values need to be evaluated as a measure of confidence in the decoded bit. For a general estimator, the LLRs of any symbol constellation with equiprobable symbols can be written as [64]

$$\Lambda(b_{ki}|\hat{x}_i) = \log\frac{\Pr(b_{ki}=1|\hat{x}_i)}{\Pr(b_{ki}=0|\hat{x}_i)} = \log\frac{\sum\limits_{q\in S(b_{ki}=1)} p(\hat{x}_i|s^{(q)})}{\sum\limits_{q\in S(b_{ki}=0)} p(\hat{x}_i|s^{(q)})}, \tag{4.195}$$

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} symbol of such a set. Further, these symbols usually have zero mean. In (4.195), $p(\hat{x}_i|s^{(q)})$ denotes the conditional PDF of the estimate $\hat{x}_i$ given that the actual transmitted symbol was $s^{(q)}$. Due to central limit theorem arguments, $p(\hat{x}_i|s^{(q)})$ can be well approximated as Gaussian for long enough data vectors. Its complex proper Gaussian approximation is determined by the conditional mean and the conditional variance according to

$$p(\hat{x}_i|s^{(q)}) = \frac{1}{\pi\mathrm{var}(\hat{x}_i|s^{(q)})} e^{-\frac{1}{\mathrm{var}(\hat{x}_i|s^{(q)})}\left|\hat{x}_i - E[\hat{x}_i|s^{(q)}]\right|^2}. \tag{4.196}$$

Together with (4.195), the LLRs of any linear estimator can be evaluated by inserting the conditional mean and the conditional variance of the specific estimator. This is now executed for the CWCU LMMSE estimator for which we obtain

$$p(\hat{x}_{\mathrm{CL},i}|s^{(q)}) = \frac{1}{\pi\mathrm{var}(\hat{x}_{\mathrm{CL},i}|s^{(q)})} \mathrm{e}^{-\frac{1}{\mathrm{var}(\hat{x}_{\mathrm{CL},i}|s^{(q)})}\left|\hat{x}_{\mathrm{CL},i} - E[\hat{x}_{\mathrm{CL},i}|s^{(q)}]\right|^2}. \tag{4.197}$$

Utilizing (4.165) and (4.166), we obtain

$$\begin{aligned}
p(\hat{x}_{\mathrm{CL},i}|s^{(q)}) &= \frac{1}{\pi[\mathbf{D}]_{i,i}^2\mathrm{var}(\hat{x}_{\mathrm{L},i}|s^{(q)})} \mathrm{e}^{-\frac{1}{[\mathbf{D}]_{i,i}^2\mathrm{var}(\hat{x}_{\mathrm{L},i}|s^{(q)})}\left|[\mathbf{D}]_{i,i}\left(\hat{x}_{\mathrm{L},i} - [\mathbf{D}]_{i,i}^{-1}s^{(q)}\right)\right|^2} \\
&= \frac{1}{\pi[\mathbf{D}]_{i,i}^2\mathrm{var}(\hat{x}_{\mathrm{L},i}|s^{(q)})} \mathrm{e}^{-\frac{1}{\mathrm{var}(\hat{x}_{\mathrm{L},i}|s^{(q)})}\left|\hat{x}_{\mathrm{L},i} - E[\hat{x}_{\mathrm{L},i}|s^{(q)}]\right|^2} \\
&= [\mathbf{D}]_{i,i}^{-2}p(\hat{x}_{\mathrm{L},i}|s^{(q)}).
\end{aligned} \tag{4.198}$$

This holds for any symbol $s^{(q)}$. The constant scaling factor $[\mathbf{D}]_{i,i}^{-2}$ does not depend on the symbol $s^{(q)}$ and it appears in the numerator and the denominator of (4.195). Thus, it cancels out, and the LLRs of the CWCU LMMSE estimates and the LLRs of the LMMSE estimates are equal for proper constellation diagrams. Therefore, also the resulting BERs of the LMMSE and the CWCU LMMSE estimators are the same, even though the BMSE of the LMMSE estimator is in general lower than that of the CWCU LMMSE estimator.

Another simulation example, taken again from the field of wireless communications, is presented in the following.

Example 4.4 (Channel Impulse Response Estimation)

As an application to demonstrate the properties of the CWCU LMMSE estimator we choose the well-known channel estimation problem for IEEE 802.11a/g/n wireless local area network (WLAN) standards [65]. The standards are based on the orthogonal frequency division multiplexing (OFDM) technology. In practice the channel frequency response estimation is of essential importance in this application. In addition, we will also discuss the channel impulse response (CIR) estimation in this example. This will particularly demonstrate the nice properties of the investigated CWCU LMMSE estimator.

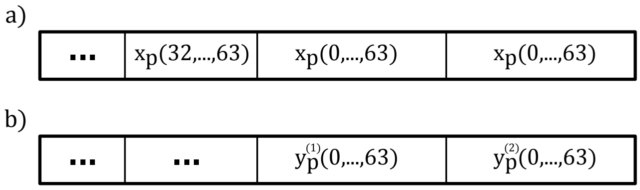

Figure 4.3: Schematic visualization of parts of the time-domain transmit and receive vectors in OFDM communications. a) Preamble including two long training symbols and a long guard interval for channel estimation; b) received long training symbols.

Model

The regarded OFDM scheme uses an inverse fast Fourier transform (IFFT) of size $N = 64$. In total, 52 subcarriers are occupied for data or pilot transmission, and the remaining 12 subcarriers are unused (or loaded with zeros). The IEEE standard also defines a preamble, cf. Figure 4.3a. We consider the two so-called long training symbols since those are designed for channel estimation. Here $\mathbf{x}_p \in \mathbb{C}^{64}$ is a known pilot vector, which is designed such that its frequency domain version $\tilde{\mathbf{x}}_p = \mathbf{F}_N \mathbf{x}_p$ shows ± 1 at the 52 occupied subcarriers (at indices out of the set $S_1 = \{1, ..., 26, 38, ...63\}$), and zeros at the unused subcarriers (at indices out of the set $S_2 = \{0, 27, ..., 37\}$). Here, $\mathbf{F}_N$ is the DFT matrix of length $N = 64$, and $\tilde{(\cdot)}$ denotes a vector in the frequency domain. Together with the 64×52 carrier selection matrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{0}^{1\times 26} & \mathbf{0}^{1\times 26} \\ \mathbf{I}^{26\times 26} & \mathbf{0}^{26\times 26} \\ \mathbf{0}^{11\times 26} & \mathbf{0}^{11\times 26} \\ \mathbf{0}^{26\times 26} & \mathbf{I}^{26\times 26} \end{bmatrix}, \tag{4.199}$$

the vector of used (non-zero) subcarrier pilot symbols can be written as $\tilde{\mathbf{x}}_{p,u} = \mathbf{B}^T \mathbf{F}_N \mathbf{x}_p$. $\mathbf{B}^T$ basically deletes the elements of the frequency domain vector $\tilde{\mathbf{x}}_p$ that correspond to the unused subcarriers. We furthermore introduce the diagonal matrix

$\mathbf{D}_p = \text{diag}\{\tilde{\mathbf{x}}_{p,u}\}$, which fulfills $\mathbf{D}_p^H \mathbf{D}_p = \mathbf{I}$ because of $\tilde{\mathbf{x}}_{p,u} \in \{-1, 1\}^{52}$. This identity will be required for deriving the estimators below.

The CIR is modeled as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{hh}})$, with

$$\mathbf{C}_{\mathbf{hh}} = \text{diag}\{\sigma_0^2, \sigma_1^2, ..., \sigma_{N_{\mathbf{h}}-1}^2\}, \tag{4.200}$$

and an exponentially decaying power delay profile according to

$$\sigma_i^2 = \Big(1 - \exp(-\tfrac{T_s}{\tau_{rms}})\Big)\exp(-\tfrac{iT_s}{\tau_{rms}}) \quad i = 0, 1, ..., N_{\mathbf{h}} - 1. \tag{4.201}$$

Here, $N_{\mathbf{h}}$ is the length of the CIR in time domain. Further, T_s and τ_{rms} are the sampling time and the channel delay spread, respectively. These two parameters are chosen as $T_s = 50\,\text{ns}$ and $\tau_{rms} = 100\,\text{ns}$ in our setup. Note that the channel length $N_{\mathbf{h}}$ can be assumed to be considerably smaller than the FFT length N. In the following we assume $N_{\mathbf{h}} = 16$.

Let $\mathbf{y}_p^{(1)}$ and $\mathbf{y}_p^{(2)}$ be the two received, channel distorted time domain preamble symbols, cf. Figure 4.3b, $\tilde{\mathbf{y}}_{p,u}^{(i)} = \mathbf{B}^T \mathbf{F}_N \mathbf{y}_p^{(i)}$ for $i = 1, 2$, and $\bar{\tilde{\mathbf{y}}} = \frac{1}{2}(\tilde{\mathbf{y}}_{p,u}^{(1)} + \tilde{\mathbf{y}}_{p,u}^{(2)})$. Then $\bar{\tilde{\mathbf{y}}}$ can be modeled as

$$\bar{\tilde{\mathbf{y}}} = \mathbf{D}_p \tilde{\mathbf{h}}_u + \tilde{\mathbf{n}} \tag{4.202}$$
$$= \mathbf{D}_p \mathbf{B}^T \tilde{\mathbf{h}} + \tilde{\mathbf{n}} \tag{4.203}$$
$$= \underbrace{\mathbf{D}_p \mathbf{B}^T \mathbf{M}_1}_{\mathbf{H}} \mathbf{h} + \tilde{\mathbf{n}} \tag{4.204}$$
$$= \mathbf{H}\mathbf{h} + \tilde{\mathbf{n}}. \tag{4.205}$$

Here $\tilde{\mathbf{h}}_u \in \mathbb{C}^{52}$ is the frequency response at the used subcarriers, $\tilde{\mathbf{h}} \in \mathbb{C}^{64}$ is the full-length frequency response including the unused frequency bins, and $\tilde{\mathbf{n}}$ is a zero mean complex proper Gaussian noise vector with covariance matrix $\mathbf{C}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} = (N\sigma_n^2/2)\mathbf{I}$, where σ_n^2 is the time domain noise variance. $\mathbf{M}_1 \in \mathbb{C}^{64\times16}$ consists of the first $N_{\mathbf{h}}$ columns of $\mathbf{F}_N$.

With (4.205) , the problem at hand has been expressed in a way such that the BLUE [45], the LMMSE estimator and the CWCU LMMSE estimator can be applied. The results are discussed in the following.

Performance Discussion

Figure 4.4 shows the BMSEs of the estimated CIR coefficients for the different estimators for the particular choice of $\sigma_n^2 = 0.01$. It is seen that the BLUE performs miserable, while the CWCU LMMSE estimator and the LMMSE estimator show a significantly better performance. The poor performance of the BLUE mainly originates from the fact that measurements are only available at the 52 frequency positions with indices out of S_1. We considered the knowledge of the impulse response duration when deriving the BLUE, however, the lack of information at the subcarriers with indices out of S_2 does not allow to reconstruct the impulse response with low MSEs. The CWCU LMMSE estimator and the LMMSE estimator use the additional prior knowledge from (4.200). As a consequence the CWCU LMMSE estimator significantly outperforms the BLUE. Furthermore, it is not far behind the LMMSE estimator, and in contrast to the LMMSE estimator it additionally shows the beneficial property of conditional unbiasedness.

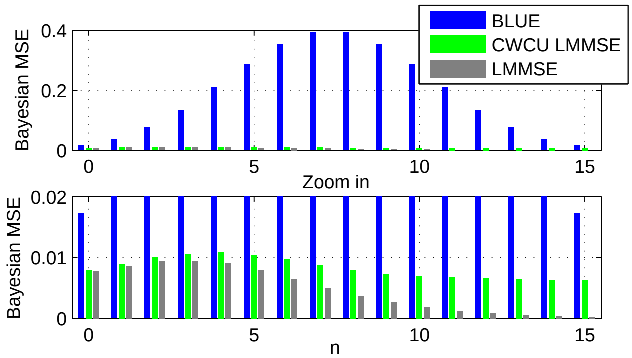

Figure 4.4: Bayesian MSEs of the estimated CIR coefficients.

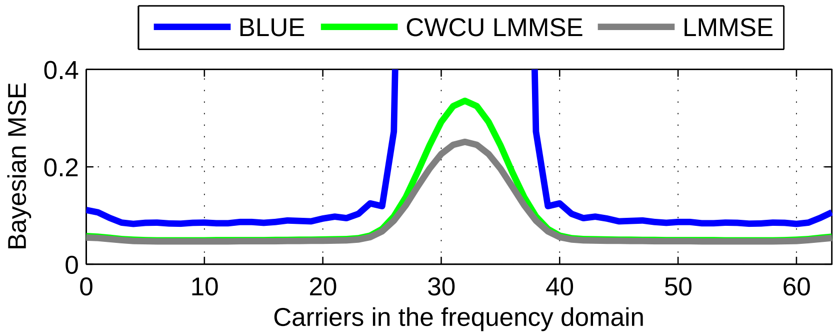

Figure 4.5: Bayesian MSE for the elements of $\hat{\tilde{\mathbf{h}}}_{\mathrm{B}}$, $\hat{\tilde{\mathbf{h}}}_{\mathrm{L}}$, and $\hat{\tilde{\mathbf{h}}}_{\mathrm{CL}}$, respectively.

In order to analyze the estimates in more detail, and in particular to explain the poor performance of the BLUE, the corresponding frequency response estimators are reviewed in the following. The LMMSE estimator $\hat{\tilde{\mathbf{h}}}_{\mathrm{L}}$ is simply obtained by computing the DFT of $\begin{bmatrix} \hat{\mathbf{h}}_{\mathrm{L}}^T & \mathbf{0}^T \end{bmatrix}^T$ (since it commutes over linear transformations). The BLUE $\hat{\tilde{\mathbf{h}}}_{\mathrm{B}}$ can be derived correspondingly since Result 3.1 can be applied. Differently, the CWCU LMMSE estimator $\hat{\tilde{\mathbf{h}}}_{\mathrm{CL}}$ cannot be derived in this way since it does not commute over general linear transformations. Also note that the vector of frequency response coefficients $\tilde{\mathbf{h}} \in \mathbb{C}^{64}$ (which corresponds to the DFT of the zero-padded impulse response $\begin{bmatrix} \mathbf{h}^T & \mathbf{0}^T \end{bmatrix}^T$) consists of complex proper Gaussian elements. Still, the PDF of $\tilde{\mathbf{h}}$ cannot be written in the form of a multivariate complex proper Gaussian

PDF. However, although $\tilde{\mathbf{h}}$ is not a complex proper Gaussian vector, $E_{\bar{\tilde{\mathbf{h}}}_i|\tilde{h}_i}[\bar{\tilde{\mathbf{h}}}_i|\tilde{h}_i]$ (with $\tilde{h}_i$ being the i^{th} element of $\tilde{\mathbf{h}}$ and with $\bar{\tilde{\mathbf{h}}}_i$ being $\tilde{\mathbf{h}}$ without $\tilde{h}_i$) is linear in $\tilde{h}_i$ (for all $i = 0, 1, \cdots, N-1$),. Therewith, one can easily show that (4.94)–(4.97) can be applied to determine the CWCU LMMSE estimator.

Figure 4.5 shows the Bayesian MSEs of $\hat{\tilde{\mathbf{h}}}_{\text{B}}$, $\hat{\tilde{\mathbf{h}}}_{\text{L}}$, and $\hat{\tilde{\mathbf{h}}}_{\text{CL}}$, respectively. $\hat{\tilde{\mathbf{h}}}_{\text{B}}$ is outperformed by $\hat{\tilde{\mathbf{h}}}_{\text{L}}$ and $\hat{\tilde{\mathbf{h}}}_{\text{CL}}$ at all frequencies, but the performance loss is significant at the large gap from subcarrier 27 to 37, where no training information is available. In contrast, $\hat{\tilde{\mathbf{h}}}_{\text{L}}$ and $\hat{\tilde{\mathbf{h}}}_{\text{CL}}$ show excellent interpolation properties along this gap. Large estimation errors of $\hat{\tilde{\mathbf{h}}}_{\text{B}}$ in this spectral region are spread over all time domain samples, which explains the poor performance of $\hat{\mathbf{h}}_{\text{B}}$. Note that in practice this is only critical if $\hat{\mathbf{h}}_{\text{B}}$ is incorporated in the receiver processing. Anyhow, pure frequency domain receivers only require estimates at the occupied 52 subcarrier positions.

4.3 Widely Linear CWCU Estimation

The intent of this section is to extend the theoretical framework of CWCU linear estimation to CWCU widely linear estimators. These investigations will lead to the CWCU WLMMSE estimator [61]. The CWCU WLMMSE estimator will be compared with the BWLUE and the WLMMSE estimator. From the previous investigations it is clear that for the LMMSE and WLMMSE estimators the particular form of the joint PDF $p(\mathbf{y}, \mathbf{x})$ does not play a role. In fact, these estimators are unambiguously defined by their first and second order statistics. As for linear CWCU estimators, this is not the case for widely linear CWCU estimators. Thus, we investigate model assumptions that allow finding 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

1. under the assumption of jointly generalized complex Gaussian $\mathbf{x}$ and $\mathbf{y}$,
2. under the linear model assumption with generalized complex Gaussian $\mathbf{x}$ and zero mean noise with known second order statistics,
3. under the linear model assumption with mutually independent complex (and otherwise arbitrarily distributed) parameters and zero mean noise with known second order statistics,
4. under the assumption of real $\mathbf{x}$, complex $\mathbf{y}$, and jointly Gaussian $\mathbf{x}$, $\text{Re}\{\mathbf{y}\}$, and $\text{Im}\{\mathbf{y}\}$,
5. under the linear model assumption with real Gaussian $\mathbf{x}$ and zero mean complex noise with known second order statistics, and
6. under the linear model assumption with mutually independent real (and other-

wise arbitrarily distributed) parameters and zero mean complex noise with known second order statistics.

We distinguish real and complex-valued parameters for reasons that will become clear soon.

4.3.1 Complex Parameter Vectors

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

$$\hat{x}_i = \mathbf{f}_i^H \mathbf{y} + \mathbf{g}_i^H \mathbf{y}^* + b_i, \quad \text{for } i = 1, 2, ..., n, \tag{4.206}$$

which can also be written as

$$\hat{x}_i = \mathbf{e}_i^H \underline{\mathbf{y}} + b_i, \quad \text{for } i = 1, 2, ..., n \tag{4.207}$$

when using

$$\mathbf{e}_i^H = \begin{bmatrix} \mathbf{f}_i^H & \mathbf{g}_i^H \end{bmatrix}. \tag{4.208}$$

The conditional mean of the estimator in (4.207) becomes

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] + b_i. \tag{4.209}$$

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 $\underline{\mathbf{x}}_i = [x_i \quad x_i^*]^T$, specifically

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]). \tag{4.210}$$

This leads to

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H \left(E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])\right) + b_i. \tag{4.211}$$

By setting (4.211) equal to $x_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{x}}_i$ we find that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{4.212}$$

$$b_i = E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. \tag{4.213}$$

These are the two conditions the widely linear estimator in (4.207) 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 becomes

$$
\begin{aligned}
J(\mathbf{e}_i) =& E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2] && (4.214)\\
=& E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \underline{\mathbf{y}} + b_i - x_i|^2] && (4.215)\\
=& E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H (\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - (x_i - E_{x_i}[x_i])|^2] && (4.216)\\
=& E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H (\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - \begin{bmatrix}1 & 0\end{bmatrix}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])|^2] && (4.217)\\
=& \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix}1\\0\end{bmatrix} - \\
& \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \mathbf{e}_i + \underbrace{\begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix}1\\0\end{bmatrix}}_{\sigma_{x_i}^2}. && (4.218)
\end{aligned}
$$

This result can be simplified by using (4.212), leading to the final optimization problem

$$
\mathbf{e}_{\mathrm{CWL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2\right) \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix}1 & 0\end{bmatrix}, \qquad (4.219)
$$

where the subscript CWL indicates the CWCU WLMMSE estimator, which will now be solved using the Lagrange multiplier method. The Lagrangian cost function is given by

$$
\begin{aligned}
\mathcal{L}(\mathbf{e}_i) = & \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2 + \left(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i}\right) \boldsymbol{\lambda} \\
& + \left(\mathbf{e}_i^T \underline{\mathbf{C}}_{\mathbf{y}x_i}^* - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i}^*\right) \boldsymbol{\lambda}^*. && (4.220)
\end{aligned}
$$

Setting the Wirtinger derivative of (4.220) w.r.t. $\mathbf{e}_i$ equal to zero allows

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} &= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} + \boldsymbol{\lambda}^H \underline{\mathbf{C}}_{x_i\mathbf{y}} \stackrel{!}{=} 0 && (4.221)\\
\mathbf{e}_{\mathrm{CWL},i}^H &= -\boldsymbol{\lambda}^H \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. && (4.222)
\end{aligned}
$$

Inserting this result into the constraint in (4.219) yields

$$
\begin{aligned}
-\boldsymbol{\lambda}^H \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i} &= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} && (4.223)\\
-\boldsymbol{\lambda}^H &= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \left(\underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1}. && (4.224)
\end{aligned}
$$

Finally, combining (4.222) and (4.224) produces

$$
\mathbf{e}_{\mathrm{CWL},i}^H = \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \left(\underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \qquad (4.225)
$$

We now denote

$$
\underline{\mathbf{D}}_i = \underline{\mathbf{C}}_{x_i x_i} \left(\underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1} \in \mathbb{C}^{2\times 2} \qquad (4.226)
$$

such that (4.225) reads as

$$
\mathbf{e}_{\mathrm{CWL},i}^H = \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{D}}_i \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \qquad (4.227)
$$

The full expression for $\hat{x}_{\mathrm{CWL},i}$ can be found by combining (4.207), (4.213) and (4.227), which yields

$$\hat{x}_{\mathrm{CWL},i} = E_{x_i}[x_i] + \mathbf{e}_{\mathrm{CWL},i}^H \big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big). \tag{4.228}$$

Using

$$\mathbf{E}_{\mathrm{CWL}} = \begin{bmatrix} \mathbf{e}_{\mathrm{CWL},1}^H \\ \mathbf{e}_{\mathrm{CWL},2}^H \\ \vdots \\ \mathbf{e}_{\mathrm{CWL},N_{\mathbf{x}}}^H \end{bmatrix} \in \mathbb{C}^{N_{\mathbf{x}} \times 2N_{\mathbf{y}}} \tag{4.229}$$

immediately leads to the vector notation of the CWCU WLMMSE estimator

$$\hat{\mathbf{x}}_{\mathrm{CWL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}} \big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big). \tag{4.230}$$

Note the similarities between the CWCU WLMMSE estimator in (4.230) and the WLMMSE estimator in (4.52). According to (4.227), the CWCU WLMMSE estimator matrix $\mathbf{E}_{\mathrm{CWL}}$ can be derived from the augmented WLMMSE estimator matrix $\underline{\mathbf{E}}_{\mathrm{WL}} = \underline{\mathbf{C}}_{\mathbf{xy}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}$ according to

$$\mathbf{E}_{\mathrm{CWL}} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 \end{bmatrix} \underline{\mathbf{E}}_{\mathrm{WL}}, \tag{4.231}$$

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

$$[\mathbf{D}_1]_{i,i} = [\underline{\mathbf{D}}_i]_{1,1}, \tag{4.232}$$

$$[\mathbf{D}_2]_{i,i} = [\underline{\mathbf{D}}_i]_{1,2}. \tag{4.233}$$

In the following, we denote $\widetilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 \end{bmatrix}$ such that (4.231) reads as

$$\mathbf{E}_{\mathrm{CWL}} = \widetilde{\mathbf{D}} \underline{\mathbf{E}}_{\mathrm{WL}}. \tag{4.234}$$

Having derived the CWCU WLMMSE estimator, we now discuss its performance measures. We begin with the mean of the error

$$E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}_{\mathrm{CWL}} - \mathbf{x}] = E_{\mathbf{y},\mathbf{x}}\big[E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \mathbf{x}\big] \tag{4.235}$$

$$= -E_{\mathbf{x}}[\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]] + \mathbf{E}_{\mathrm{CWL}} E_{\mathbf{y}}\big[\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big] \tag{4.236}$$

$$= \mathbf{0}. \tag{4.237}$$

With that, the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{CWL}}$ follows as

$$\mathbf{C}_{\mathbf{ee},\mathrm{CWL}} = E_{\mathbf{y},\mathbf{x}}\big[\mathbf{e}\mathbf{e}^H\big] \tag{4.238}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \mathbf{x}\big)\big(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \mathbf{x}\big)^H\Big] \tag{4.239}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\big) \times \big(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\big)^H\Big] \tag{4.240}$$

$$= \mathbf{C}_{\mathbf{xx}} - \mathbf{E}_{\mathrm{CWL}}\mathbf{C}_{\underline{\mathbf{y}}\mathbf{x}} - \mathbf{C}_{\mathbf{x}\underline{\mathbf{y}}}\mathbf{E}_{\mathrm{CWL}}^H + \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{E}_{\mathrm{CWL}}^H \tag{4.241}$$

$$= \mathbf{C}_{\mathbf{xx}} - \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yx}}\begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\\ \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix} - \begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}} & \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix}\underline{\mathbf{C}}_{\mathbf{xy}}\mathbf{E}_{\mathrm{CWL}}^H + \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{E}_{\mathrm{CWL}}^H. \tag{4.242}$$

By defining the augmented matrix $\underline{\mathbf{M}}$ as

$$\underline{\mathbf{M}} = \underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{yx}}, \tag{4.243}$$

(4.242) simplifies to

$$\mathbf{C}_{\mathbf{ee},\mathrm{CWL}} = \mathbf{C}_{\mathbf{xx}} - \widetilde{\mathbf{D}}\underline{\mathbf{M}}\begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\\ \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix} - \begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}} & \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix}\underline{\mathbf{M}}\widetilde{\mathbf{D}}^H + \widetilde{\mathbf{D}}\underline{\mathbf{M}}\widetilde{\mathbf{D}}^H. \tag{4.244}$$

The BMSE values of the i^{th} estimate $\hat{x}_{\mathrm{CWL},i}$ corresponds to the i^{th} diagonal element of the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{CWL}}$.

The findings of this section so far lead to case 1 of

Result 4.2 (CWCU WLMMSE Estimator for Complex-Valued Parameter Vectors)

If $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued parameter vector and

1. $\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^m$ are generalized jointly Gaussian, or
2. $\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^m$ are connected via the linear model in (4.1) 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
3. $\mathbf{x}$ and $\mathbf{y} \in \mathbb{C}^m$ are connected via the linear model in (4.1) and $\mathbf{x}$ has mean $E_{\mathbf{x}}[\mathbf{x}]$ and mutually independent elements such that $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \cdots, \sigma_{x_{N_\mathbf{x}}}^2\}$ and $\widetilde{\mathbf{C}}_{\mathbf{xx}} = \mathrm{diag}\{\widetilde{\sigma}_{x_1}^2, \widetilde{\sigma}_{x_2}^2, \cdots, \widetilde{\sigma}_{x_{N_\mathbf{x}}}^2\}$ (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 $i = 1, 2, \cdots, N_\mathbf{x}$ is given by (4.230), where the estimator matrix $\mathbf{E}_{\mathrm{CWL}}$ is defined in (4.231)–(4.233) and (4.226). The mean of the error $\mathbf{e} = \hat{\mathbf{x}}_{\mathrm{CWL}} - \mathbf{x}$ (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 provided in (4.244) with $\underline{\mathbf{M}}$ defined in (4.243). The minimum BMSEs are $\mathrm{Bmse}(\hat{x}_{\mathrm{CWL},i}) = [\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CWL}}}]_{i,i}$.

If none of the three cases is fulfilled, then in the linear model setup a widely linear CWCU estimator is available in form of the BWLUE, which not necessarily has to correspond to the CWCU WLMMSE estimator.

Case 2 and 3 in Result 4.2 (derived in Appendix S) originate from similar considerations as for case 2 and 3 in Result 4.1. Again, a significant relaxation of the jointly Gaussian assumption for $\mathbf{x}$ and $\mathbf{y}$ can be achieved by incorporating the linear model assumption already earlier in the derivation of the estimator. Note that in a linear model setup the required statistics become

$$\underline{\mathbf{C}}_{x_i\mathbf{y}} = \underline{\mathbf{C}}_{x_i\mathbf{x}}\underline{\mathbf{H}}^H \tag{4.245}$$

$$\underline{\mathbf{C}}_{\mathbf{y}x_i} = \underline{\mathbf{H}}\,\underline{\mathbf{C}}_{\mathbf{x}x_i} \tag{4.246}$$

$$\underline{\mathbf{C}}_{\mathbf{yy}} = \underline{\mathbf{H}}\,\underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H + \underline{\mathbf{C}}_{\mathbf{nn}} \tag{4.247}$$

$$\underline{\mathbf{C}}_{\mathbf{xy}} = \underline{\mathbf{C}}_{\mathbf{xx}}\underline{\mathbf{H}}^H. \tag{4.248}$$

Having derived the CWCU WLMMSE estimator, some further details are investigated.

4.3.2 Discussion of the CWCU WLMMSE Estimator for Complex-Valued Parameters

Commonalities between the Three Cases in Result 4.2

For all three cases in Result 4.2 it holds that

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]). \tag{4.249}$$

This is obvious for the first case where $\mathbf{y}$ and $\mathbf{x}$ are jointly Gaussian. The proof for the other two cases is provided in Appendix T.

Similarly, it holds for all three cases that the augmented conditional covariance matrix $\underline{\mathbf{C}}_{\mathbf{yy}|x_i}$ is given by

$$\underline{\mathbf{C}}_{\mathbf{yy}|x_i} = \underline{\mathbf{C}}_{\mathbf{yy}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}. \tag{4.250}$$

The proof is presented in Appendix U.

Conditional Properties

In the following, the BWLUE, the WLMMSE estimator and the CWCU WLMMSE estimator are analyzed in terms of their conditional mean $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i]$, conditional bias $b(\hat{x}_i|x_i)$, conditional variance $\mathrm{var}(\hat{x}_i|x_i)$ and conditional MSE $\mathrm{mse}(\hat{x}_i|x_i)$.

We begin with the BWLUE, which will be analyzed from a Bayesian perspective. This is valid since one can show that the BWLUE can also be derived by minimizing the BMSE cost function subject to an unbiased constraint in a similar manner as it was done for the BLUE in Appendix J. Consider the BWLUE for x_i in (3.62) for which we obtain the conditional properties

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{BW},i}|x_i] = x_i, \tag{4.251}$$
$$b(\hat{x}_{\mathrm{BW},i}|x_i) = 0, \tag{4.252}$$
$$\mathrm{var}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\mathbf{u}_i, \tag{4.253}$$
$$\mathrm{mse}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\mathbf{u}_i. \tag{4.254}$$

The derivation of (4.251)–(4.254) can be found in Appendix V. Note that $\mathrm{var}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{BW},i})$ and $\mathrm{mse}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathrm{mse}(\hat{x}_{\mathrm{BW},i})$ hold for the BWLUE.

For the derivation of the corresponding properties for the WLMMSE estimator we assume that at least one of the three cases mentioned in Result 4.2 holds. Then, the following conditional properties are obtained

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] = \begin{bmatrix} 1 & 0 \end{bmatrix} \big(\underline{\mathbf{D}}_i^{-1}\underline{\mathbf{x}}_i + (\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}) E_{x_i}[\underline{\mathbf{x}}_i]\big), \tag{4.255}$$

$$b(\hat{x}_{\mathrm{WL},i}|x_i) = \begin{bmatrix} 1 & 0 \end{bmatrix} \big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]), \tag{4.256}$$

$$\mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{4.257}$$

$$\begin{aligned} \mathrm{mse}(\hat{x}_{\mathrm{WL},i}|x_i) = {} & \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ & + \left| \begin{bmatrix} 1 & 0 \end{bmatrix} \big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) \right|^2. \end{aligned} \tag{4.258}$$

For the derivations we refer to Appendix W.

The conditional properties in (4.251)–(4.254) for the BWLUE and in (4.255)–(4.258) for the WLMMSE estimator are now compared with those for the CWCU WLMMSE estimator, which according to Appendix X are given by

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] = x_i, \tag{4.259}$$

$$b(\hat{x}_{\mathrm{CWL},i}|x_i) = 0, \tag{4.260}$$

$$\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{D}}_i \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \sigma_{x_i}^2, \tag{4.261}$$

$$\mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i) = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{D}}_i \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \sigma_{x_i}^2. \tag{4.262}$$

This comparison leads to following statements, which are very similar to the corresponding statements for the linear estimators in Section 4.2.3: The conditional mean and the conditional bias of the CWCU WLMMSE estimator correspond to those of the BWLUE. The WLMMSE estimator is conditionally biased as it can be seen in (4.256). For the special case of a zero mean parameter $E_{x_i}[x_i] = 0$, the augmented conditional means of the WLMMSE and CWCU WLMMSE estimators read as

$$E_{\mathbf{y}|x_i}[\hat{\underline{\mathbf{x}}}_{\mathrm{WL},i}|x_i] = \underline{\mathbf{D}}_i^{-1}\underline{\mathbf{x}}_i \tag{4.263}$$

and

$$E_{\mathbf{y}|x_i}[\hat{\underline{\mathbf{x}}}_{\mathrm{CWL},i}|x_i] = \underline{\mathbf{x}}_i, \tag{4.264}$$

respectively. This directly leads to

$$E_{\mathbf{y}|x_i}[\hat{\underline{\mathbf{x}}}_{\mathrm{CWL},i}|x_i] = \underline{\mathbf{D}}_i E_{\mathbf{y}|x_i}[\hat{\underline{\mathbf{x}}}_{\mathrm{WL},i}|x_i]. \tag{4.265}$$

Note that for this special case

$$\hat{\underline{\mathbf{x}}}_{\mathrm{CWL},i} = \underline{\mathbf{D}}_i \hat{\underline{\mathbf{x}}}_{\mathrm{WL},i} \tag{4.266}$$

holds.

From (4.257) the augmented conditional covariance matrix of $\hat{x}_{\mathrm{WL},i}$ conditioned on x_i directly follows as

$$\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{WL}} = \underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_ix_i}. \tag{4.267}$$

The corresponding expression for the CWCU WLMMSE estimator follows from (X.10) in Appendix X to be

$$\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{CWL}} = \underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_ix_i} - \underline{\mathbf{C}}_{x_ix_i}. \tag{4.268}$$

Combining (4.267) and (4.268) allows to identify the relation

$$\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{CWL}} = \underline{\mathbf{D}}_i^2\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{WL}}. \tag{4.269}$$

Another notation of (4.269) that turns out to be useful later is

$$\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{CWL}} = \underline{\mathbf{D}}_i\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{WL}}\underline{\mathbf{D}}_i^H, \tag{4.270}$$

which can be proven the following way

$$\begin{aligned}
\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{WL}}\underline{\mathbf{D}}_i^H &= \underline{\mathbf{D}}_i\underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{D}}_i^H && (4.271)\\
&= \underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{D}}_i^H - \underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{D}}_i^H && (4.272)\\
&= \underline{\mathbf{C}}_{x_ix_i}\big(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\big)^{-1}\underline{\mathbf{C}}_{x_ix_i}\\
&\quad - \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_ix_i}\big(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\big)^{-1}\underline{\mathbf{C}}_{x_ix_i} && (4.273)\\
&= \underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_ix_i} - \underline{\mathbf{C}}_{x_ix_i} && (4.274)\\
&= \underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|x_i,\mathrm{CWL}}. && (4.275)
\end{aligned}$$

As for the linear case, $\mathrm{mse}(\hat{x}_{\mathrm{WL},i}|x_i)$ clearly depends on the actual realization of x_i, while $\mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i)$ does not. However, by averaging the conditional MSE over the PDF of x_i, the BMSE for the WLMMSE estimator follows as

$$\begin{aligned}
\mathrm{Bmse}(\hat{x}_{\mathrm{WL},i}) &= E_{x_i}[\mathrm{mse}(\hat{x}_{\mathrm{WL},i}|x_i)] && (4.276)\\
&= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}\\
&\quad + \begin{bmatrix}1 & 0\end{bmatrix}\big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)\underline{\mathbf{C}}_{x_ix_i}\big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)^H\begin{bmatrix}1\\0\end{bmatrix} && (4.277)\\
&= \begin{bmatrix}1 & 0\end{bmatrix}\Big(\underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i} - \big(\underline{\mathbf{D}}_i^{-1}\big)^2\underline{\mathbf{C}}_{x_ix_i} + \underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i}\big(\underline{\mathbf{D}}_i^{-1}\big)^H\\
&\quad - \underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i} - \underline{\mathbf{C}}_{x_ix_i}\big(\underline{\mathbf{D}}_i^{-1}\big)^H + \underline{\mathbf{C}}_{x_ix_i}\Big)\begin{bmatrix}1\\0\end{bmatrix} && (4.278)\\
&= \begin{bmatrix}1 & 0\end{bmatrix}\Big(\underline{\mathbf{C}}_{x_ix_i} - \underline{\mathbf{C}}_{x_ix_i}\big(\underline{\mathbf{D}}_i^{-1}\big)^H\Big)\begin{bmatrix}1\\0\end{bmatrix} && (4.279)\\
&= \sigma_{x_i}^2 - \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}. && (4.280)
\end{aligned}$$

Interestingly, the terms $-\left(\underline{\mathbf{D}}_i^{-1}\right)^2\underline{\mathbf{C}}_{x_ix_i}$ and $\underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{D}}_i^{-1}\right)^H$ in (4.278) cancel each other since

$$\underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{D}}_i^{-1}\right)^H = \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\right)^H \tag{4.281}$$

$$= \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i} \tag{4.282}$$

$$= \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_ix_i} \tag{4.283}$$

$$= \left(\underline{\mathbf{D}}_i^{-1}\right)^2\underline{\mathbf{C}}_{x_ix_i}. \tag{4.284}$$

For the CWCU WLMMSE estimator we trivially obtain

$$\mathrm{Bmse}(\hat{x}_{\mathrm{CWL},i}) = \mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i) = \begin{bmatrix} 1 & 0 \end{bmatrix}\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \sigma_{x_i}^2. \tag{4.285}$$

For the BWLUE we directly obtain $\mathrm{mse}(\hat{x}_{\mathrm{WB},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{WB},i}|x_i) = \mathrm{Bmse}(\hat{x}_{\mathrm{WB},i})$.

Relation to the BWLUE, WLMMSE Estimator and CWCU LMMSE Estimator

Similar to their linear equivalents, the CWCU WLMMSE estimator turns out to be closely related to the WLMMSE estimator. According to (4.234), the CWCU WLMMSE estimator matrix $\mathbf{E}_{\mathrm{CWL}}$ can be derived by multiplying $\underline{\mathbf{E}}_{\mathrm{WL}}$ with a matrix containing two diagonal blocks. This can be interpreted as the widely linear extension of the connection between the CWCU LMMSE estimator and the LMMSE estimator. For the special case of zero mean parameters ($E_{\mathbf{x}}[\mathbf{x}] = \mathbf{0}$), the CWCU WLMMSE estimates are widely linear transformed WLMMSE estimates as

$$\hat{\mathbf{x}}_{\mathrm{CWL}} = \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{y}} = \widetilde{\mathbf{D}}\underline{\mathbf{E}}_{\mathrm{WL}}\underline{\mathbf{y}} = \widetilde{\mathbf{D}}\hat{\underline{\mathbf{x}}}_{\mathrm{WL}}. \tag{4.286}$$

Note that the WLMMSE estimator corresponds to the LMMSE estimator when $\mathbf{x}$ and $\mathbf{n}$ are both proper. If this is the case, $\underline{\mathbf{C}}_{x_ix_i}$ and $\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}$ are both diagonal matrices of size 2×2. As a consequence, $\underline{\mathbf{D}}_i$ in (4.226) is also a diagonal matrix and the CWCU WLMMSE estimator corresponds to the CWCU LMMSE estimator.

The relation between the CWCU WLMMSE estimator and the BWLUE is similar to the relation between the CWCU LMMSE estimator and the BLUE discussed in Section 4.2.3. Hence, the CWCU WLMMSE coincides with the BWLUE when there is only one parameter to be estimated, i.e., $N_{\mathbf{x}} = 1$. Another case is when $\mathbf{C}_{\mathbf{xx}}$, $\widetilde{\mathbf{C}}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{nn}}$, $\widetilde{\mathbf{C}}_{\mathbf{nn}}$ and $\mathbf{H}$ are all diagonal matrices. The proof is provided in Appendix Y. And as already said in Result 4.2, if non of the 3 mentioned cases is fulfilled, then in the linear model setup a widely linear CWCU estimator is available in form of the BWLUE, which not necessarily has to correspond to the CWCU WLMMSE estimator.

Transformation Analysis

The CWCU WLMMSE estimator will in general not commute over affine transformations of the form

$$\boldsymbol{\alpha} = \mathbf{B}\mathbf{x} + \mathbf{c} \tag{4.287}$$

as shown in the following. Let the i^{th} row of the transformation matrix $\mathbf{B}$ be denoted by $\mathbf{b}_i^H$ such that

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^H \\ \mathbf{b}_2^H \\ \vdots \\ \mathbf{b}_{N_{\boldsymbol{\alpha}}}^H \end{bmatrix} \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}, \tag{4.288}$$

and let α_i and c_i be the i^{th} elements of $\boldsymbol{\alpha}$ and $\mathbf{c}$, respectively, such that $\alpha_i = \mathbf{b}_i^H \mathbf{x} + c_i$. The augmented versions of α_i shall be given as

$$\underline{\boldsymbol{\alpha}}_i = \underline{\mathbf{B}}_i \underline{\mathbf{x}} + \underline{\mathbf{c}}_i, \tag{4.289}$$

where

$$\underline{\mathbf{B}}_i = \begin{bmatrix} \mathbf{b}_i^H & \mathbf{0}^{1\times N_{\mathbf{x}}} \\ \mathbf{0}^{1\times N_{\mathbf{x}}} & \mathbf{b}_i^T \end{bmatrix}. \tag{4.290}$$

We now assume that one of the three cases in Result 4.2 holds for $\boldsymbol{\alpha}$ instead of $\mathbf{x}$, such that the CWCU WLMMSE estimator for α_i follows as

$$\hat{\alpha}_{\mathrm{CWL},i} = E_{\alpha_i}[\alpha_i] + \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{\alpha_i\alpha_i} \left(\underline{\mathbf{C}}_{\alpha_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}\alpha_i} \right)^{-1} \underline{\mathbf{C}}_{\alpha_i \mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1} \left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}] \right). \tag{4.291}$$

From (4.289), we can readily derive that

$$\begin{aligned} E_{\alpha_i}[\underline{\boldsymbol{\alpha}}_i] &= \underline{\mathbf{B}}_i E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{c}}_i, && (4.292) \\ E_{\alpha_i}[\alpha_i] &= \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{B}}_i E_{\mathbf{x}}[\underline{\mathbf{x}}] + c_i, && (4.293) \\ \underline{\mathbf{C}}_{\alpha_i\alpha_i} &= \underline{\mathbf{B}}_i \underline{\mathbf{C}}_{\mathbf{x}\mathbf{x}} \underline{\mathbf{B}}_i^H, && (4.294) \\ \underline{\mathbf{C}}_{\mathbf{y}\alpha_i} &= E_{\mathbf{y},\alpha_i}\left[\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}] \right) \left(\underline{\boldsymbol{\alpha}}_i - E_{\alpha_i}[\underline{\boldsymbol{\alpha}}_i] \right)^H \right], && (4.295) \\ &= E_{\mathbf{y},\mathbf{x}}\left[\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}] \right) \left(\underline{\mathbf{B}}_i (\underline{\mathbf{x}} - E_{\mathbf{x}}[\underline{\mathbf{x}}]) \right)^H \right], && (4.296) \\ &= \underline{\mathbf{C}}_{\mathbf{y}\mathbf{x}} \underline{\mathbf{B}}_i^H. && (4.297) \end{aligned}$$

Inserting (4.293)–(4.297) into the CWCU WLMMSE estimator in (4.291) directly leads to the insight that the CWCU WLMMSE estimator in general does not commute over affine transformations.

An exception can be found for diagonal transformation matrices such as

$$\mathbf{B} = \operatorname{diag}\left\{ \begin{bmatrix} b_1 & b_2 & \dots & b_{N_{\mathbf{x}}} \end{bmatrix} \right\} \in \mathbb{C}^{N_{\mathbf{x}} \times N_{\mathbf{x}}}, \tag{4.298}$$

with non-zero diagonal elements. The latter requirement corresponds to $\mathbf{B}$ being invertible. By constraining the transformation matrix to be diagonal and utilizing the notation

$$\underline{\widetilde{\mathbf{B}}}_i = \mathrm{diag}\left\{\begin{bmatrix} b_i & b_i^* \end{bmatrix}\right\} \in \mathbb{C}^{2\times 2}, \tag{4.299}$$

the expressions in (4.293)–(4.297) read as

$$\alpha_i = b_i x_i + c_i, \tag{4.300}$$
$$\underline{\boldsymbol{\alpha}}_i = \underline{\widetilde{\mathbf{B}}}_i \underline{\mathbf{x}}_i + \underline{\mathbf{c}}_i, \tag{4.301}$$
$$E_{\alpha_i}[\alpha_i] = b_i E_{x_i}[x_i] + c_i, \tag{4.302}$$
$$\underline{\mathbf{C}}_{\alpha_i\alpha_i} = \underline{\widetilde{\mathbf{B}}}_i \underline{\mathbf{C}}_{x_i x_i} \underline{\widetilde{\mathbf{B}}}_i^H, \tag{4.303}$$
$$\underline{\mathbf{C}}_{\mathbf{y}\alpha_i} = E_{\mathbf{y},\alpha_i}\left[\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right)\left(\underline{\boldsymbol{\alpha}}_i - E_{\alpha_i}[\underline{\boldsymbol{\alpha}}_i]\right)^H\right], \tag{4.304}$$
$$= E_{\mathbf{y},x_i}\left[\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right)\left(\underline{\widetilde{\mathbf{B}}}_i(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])\right)^H\right], \tag{4.305}$$
$$= \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\widetilde{\mathbf{B}}}_i^H. \tag{4.306}$$

Now, inserting (4.302)–(4.306) into the CWCU WLMMSE estimator in (4.291) and utilizing the fact that b_i is non-zero results in

$$\hat{\alpha}_{\mathrm{CWL},i} = b_i E_{x_i}[x_i] + c_i + \begin{bmatrix}1 & 0\end{bmatrix} \underline{\widetilde{\mathbf{B}}}_i \underline{\mathbf{C}}_{x_i x_i} \underline{\widetilde{\mathbf{B}}}_i^H \left(\underline{\widetilde{\mathbf{B}}}_i \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\widetilde{\mathbf{B}}}_i^H\right)^{-1} \underline{\widetilde{\mathbf{B}}}_i \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{4.307}$$
$$= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\widetilde{\mathbf{B}}}_i E_{x_i}[\underline{\mathbf{x}}_i] + c_i + \begin{bmatrix}1 & 0\end{bmatrix} \underline{\widetilde{\mathbf{B}}}_i \underline{\mathbf{C}}_{x_i x_i} \left(\underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{4.308}$$
$$= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\widetilde{\mathbf{B}}}_i \underline{\hat{\mathbf{x}}}_{\mathrm{CWL},i} + c_i \tag{4.309}$$
$$= b_i \hat{x}_{\mathrm{CWL},i} + c_i. \tag{4.310}$$

One can conclude that the CWCU WLMMSE estimator commutes over affine transformations with diagonal and invertible transformation matrices.

Example 4.5 (Log-Likelihood Ratio Evaluation of the CWCU WLMMSE Estimator)

In Example 4.3, we showed that the LLR evaluated from the estimates of the CWCU LMMSE and LMMSE estimators coincide. This was done by assuming the constellation diagram to be proper (as it is the case for QPSK symbols). We now turn to improper constellation diagrams such as quadrature amplitude modulation (QAM) with 8 symbols (8-QAM). In such scenarios it is advantageous to use widely linear estimators, which can incorporate information about the improperness of the data. Interestingly, it will turn out that also the LLRs evaluated from the estimates of the CWCU WLMMSE and WLMMSE estimators coincide [58].

Recall the definition of the LLR in (4.195), where $p(\hat{x}_i|s^{(q)})$ denotes the conditional PDF of the estimate $\hat{x}_i$ given that the actual symbol was $s^{(q)}$. Due to central limit theorem arguments, $p(\hat{x}_i|s^{(q)})$ can be well approximated as Gaussian for long enough data vectors. Its general complex Gaussian approximation is determined by the augmented conditional mean and the augmented conditional covariance matrix according to

$$p(\hat{x}_i|s^{(q)}) = \frac{1}{\sqrt{\pi^2 \det(\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)}})}} \mathrm{e}^{-\frac{1}{2}\left(\underline{\hat{\mathbf{x}}}_i - E[\underline{\hat{\mathbf{x}}}_i|s^{(q)}]\right)^H \underline{\mathbf{C}}^{-1}_{\hat{x}_i\hat{x}_i|s^{(q)}} \left(\underline{\hat{\mathbf{x}}}_i - E[\underline{\hat{\mathbf{x}}}_i|s^{(q)}]\right)}. \tag{4.311}$$

In analogy to the linear case in (4.198) it will now be shown that $p(\hat{x}_{\mathrm{WL},i}|s^{(q)})$ of the WLMMSE estimator and $p(\hat{x}_{\mathrm{CWL},i}|s^{(q)})$ of the CWCU WLMMSE estimator only differ by a constant factor. By utilizing (4.286) and (4.270) the exponent of (4.311) for the CWCU WLMMSE estimator can be rearranged to

$$-\frac{1}{2}\left(\underline{\hat{\mathbf{x}}}_{\mathrm{CWL},i} - E[\underline{\hat{\mathbf{x}}}_{\mathrm{CWL},i}|s^{(q)}]\right)^H \underline{\mathbf{C}}^{-1}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{CWL}} \left(\underline{\hat{\mathbf{x}}}_{\mathrm{CWL},i} - E[\underline{\hat{\mathbf{x}}}_{\mathrm{CWL},i}|s^{(q)}]\right) \tag{4.312}$$

$$= -\frac{1}{2}\left(\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i} - E[\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i}|s^{(q)}]\right)^H \underline{\mathbf{D}}_i^H \left(\underline{\mathbf{D}}_i \underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{WL}} \underline{\mathbf{D}}_i^H\right)^{-1}$$

$$\times \underline{\mathbf{D}}_i \left(\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i} - E[\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i}|s^{(q)}]\right) \tag{4.313}$$

$$= -\frac{1}{2}\left(\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i} - E[\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i}|s^{(q)}]\right)^H \underline{\mathbf{C}}^{-1}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{WL}} \left(\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i} - E[\underline{\hat{\mathbf{x}}}_{\mathrm{WL},i}|s^{(q)}]\right). \tag{4.314}$$

This result shows that the exponent of (4.311) is identical for the CWCU WLMMSE estimator and the WLMMSE estimator for a given $\mathbf{y}$. The prefactor of (4.311) for the CWCU WLMMSE estimator becomes

$$\frac{1}{\sqrt{\pi^2 \det(\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{CWL}})}} = \frac{1}{\sqrt{\pi^2 \det\left(\underline{\mathbf{D}}_i \underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{WL}} \underline{\mathbf{D}}_i^H\right)}} \tag{4.315}$$

$$= \frac{1}{\sqrt{\pi^2 \det(\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{WL}}) |\det(\underline{\mathbf{D}}_i)|^2}} \tag{4.316}$$

$$= \frac{\left|\det\left(\underline{\mathbf{D}}_i^{-1}\right)\right|}{\sqrt{\pi^2 \det(\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)},\mathrm{WL}})}}. \tag{4.317}$$

Like in the linear case in (4.198), the prefactors of the CWCU WLMMSE estimator and the WLMMSE estimator only differ by a constant real factor. This factor does not depend on the symbol $s^{(q)}$ and it appears in the numerator and the denominator of (4.195), thus cancelling out in the determination of the LLRs. This leads to the insight that the LLRs evaluated from the CWCU WLMMSE estimates in Result 4.2 and the WLMMSE estimates in (4.52) according to (4.195) and (4.311) are identical. As a consequence of this, the corresponding BER performance of the WLMMSE and the CWCU WLMMSE estimator coincide.

The insights of this example are beneficially applied in the following example.

Example 4.6 (Estimation of 8-QAM Symbols in a Unique Word OFDM Framework)

An example where employing the CWCU WLMMSE estimator allows 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 $S = \{-3 \pm j, -1 \pm j, 1 \pm j, 3 \pm j\}$, 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 [45,66]. Like classical OFDM, UW-OFDM is a block based transmission scheme where at the receive side a data vector $\mathbf{d}$ is estimated based 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

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{G}\mathbf{d} + \tilde{\mathbf{v}}, \tag{4.318}$$

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. [45,66]), $\mathbf{d}$ 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 Figure 4.6. The first block is implemented as a convolutional encoder with the industry standard rate 1/2, and constraint length 7 code with generator polynomials (133, 171) as defined in [65]. 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 and 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.

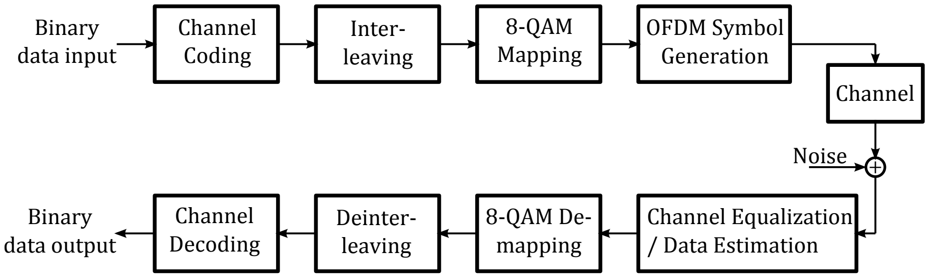

Figure 4.6: Block diagram of the investigated UW-OFDM communication system [58].

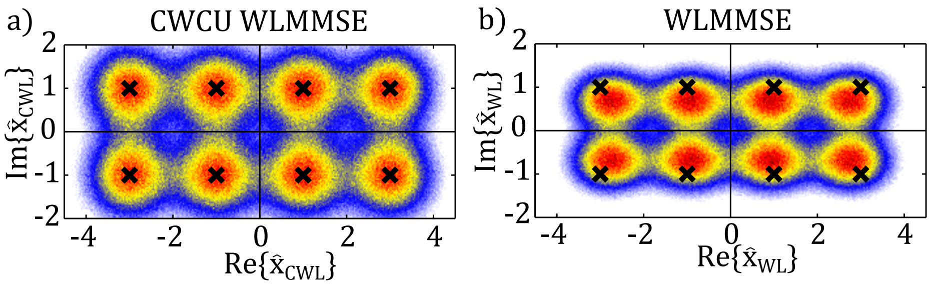

Figure 4.7: Relative frequencies of the CWCU WLMMSE estimates in (a), and the WLMMSE estimates in (b). The black crosses mark the original 8-QAM constellation points [58].

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 [45,66] has been used. 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 distribution in both cases. If the estimates are improper, the generalized complex Gaussian density function [3,23–25]

$$p(\hat{x}_i|s^{(q)}) = \frac{1}{\sqrt{\pi^2 \det(\underline{\mathbf{C}}_{\hat{x}_i\hat{x}_i|s^{(q)}})}} \mathrm{e}^{-\frac{1}{2}\left(\underline{\hat{\mathbf{x}}}_i - E[\underline{\hat{\mathbf{x}}}_i|s^{(q)}]\right)^H \underline{\mathbf{C}}^{-1}_{\hat{x}_i\hat{x}_i|s^{(q)}}\left(\underline{\hat{\mathbf{x}}}_i - E[\underline{\hat{\mathbf{x}}}_i|s^{(q)}]\right)} \tag{4.319}$$

has to be used. Otherwise the simpler complex proper Gaussian density

$$p(\hat{x}_i|s^{(q)}) = \frac{1}{\pi\sigma^2_{\hat{x}_i|s^{(q)}}} \mathrm{e}^{-\frac{1}{\sigma^2_{\hat{x}_i|s^{(q)}}}\left|\hat{x}_i - E[\hat{x}_i|s^{(q)}]\right|^2} \tag{4.320}$$

can be employed, 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 (4.319), (4.320) does not require the augmented form. Consequently, no evaluations of determinants and also no matrix inversions are required. It has been shown in [58] that for the estimated 8-QAM symbols transmitted over an 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 Figure 4.7, which is taken from [58]. 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 properly distributed, cf. Figure 4.7b. As a consequence, the CWCU WLMMSE estimator allows utilizing (4.320), while the WLMMSE estimator requires (4.319) to derive the LLRs for further processing. Furthermore, it has been shown in Example 4.5 that the LLRs and consequently the BERs of the CWCU WLMMSE estimator and the WLMMSE estimator coincide [58]. 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 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.

As further difficulty we now consider multipath channels instead of the AWGN channel used so far. The channel impulse responses (CIRs) are modeled as tapped delay lines, each tap with uniformly distributed phase and Rayleigh distributed magnitude. Further, we assume the power to decay exponentially as defined in [67]. The model allows the choice of the channel delay spread, for a more detailed description the reader may refer to [67]. In total 10 000 CIR realizations featuring a channel delay spread of $\tau_{\mathrm{RMS}} = 100\,\mathrm{ns}$ have been generated and stored, and the BER simulation results are obtained by averaging over these 10 000 realizations.

The data estimation is performed with the WLMMSE estimator, the CWCU WLMMSE estimator, as well as 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 smaller than the main diagonal elements by at least a factor of 10^{-3} in all cases. We therefore again apply (4.320) for the LLR calculation in case the CWCU WLMMSE estimator is used. The effects on the BER performance in dependence of the mean energy per bit to noise power spectral density ratio E_b/N_0 is visualized in Figure 4.8. This figure shows that the loss in performance of the CWCU WLMMSE estimator using the simplified PDF in (4.320) for LLR calculation is definitely insignificant. Note that in practice usually approximation formulas are used to derive LLRs. In our application this means that (4.320) instead of (4.319) can be used as a starting point to derive LLR approximations [68–70].

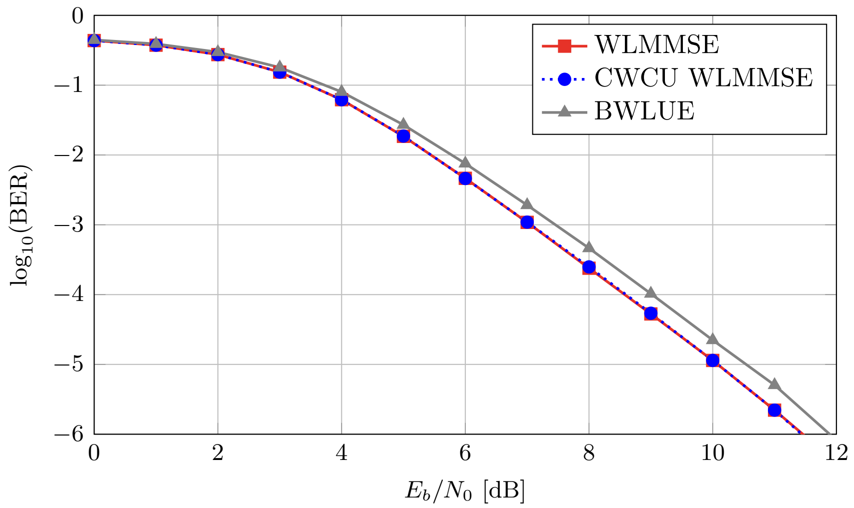

Figure 4.8: Bit error ratio of different widely linear estimators for the described digital communication system setup in a multipath scenario. For the WLMMSE estimator, (4.319) was used for the LLR determination, while for the CWCU WLMMSE estimator the simpler expression in (4.320) was applied.

4.3.3 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 $\underline{\mathbf{C}}_{x_i x_i}$ is not invertible since

$$\underline{\mathbf{C}}_{x_i x_i} = \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{x_i}^2 \\ \sigma_{x_i}^2 & \sigma_{x_i}^2 \end{bmatrix}. \tag{4.321}$$

Note that this was required in the derivation for all three cases of Result 4.2. Anyhow, we now assume the real composite vector

$$\mathbf{y}_{\mathbb{R}} = \begin{bmatrix} \mathbf{y}_{\mathrm{R}} \\ \mathbf{y}_{\mathrm{I}} \end{bmatrix} \in \mathbb{R}^{2N_{\mathbf{y}}}, \tag{4.322}$$

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

$$E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i] = E_{\mathbf{y}_{\mathbb{R}}}[\mathbf{y}_{\mathbb{R}}] + \mathbf{C}_{\mathbf{y}_{\mathbb{R}} x_i} \frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \tag{4.323}$$

By multiplying (4.323) with the real-to-complex transformation matrix $\mathbf{T}_{N_{\mathbf{y}}}$ in (2.3) from the left we obtain an expression for $E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ as

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y} x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \tag{4.324}$$

With (4.324) the conditional mean of the estimator in (4.207) becomes

$$
\begin{aligned}
E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] &= E_{\mathbf{y}|x_i}[\mathbf{e}_i^H \underline{\mathbf{y}} + b_i|x_i] && (4.325)\\
&= \mathbf{e}_i^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] + b_i && (4.326)\\
&= \mathbf{e}_i^H \left(E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\sigma_{x_i}^2} (x_i - E_{x_i}[x_i]) \right) + b_i. && (4.327)
\end{aligned}
$$

By setting (4.327) 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

$$
\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\sigma_{x_i}^2} = 1 \tag{4.328}
$$

and

$$
E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}] = b_i. \tag{4.329}
$$

In order to simplify the BMSE cost function, (4.328) and (4.329) can be used such that

$$
\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2] && (4.330)\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \underline{\mathbf{y}} + b_i - x_i|^2] && (4.331)\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H (\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - (x_i - E_{x_i}[x_i])|^2] && (4.332)\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H (\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]) - \begin{bmatrix} 1 & 0 \end{bmatrix} (\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])|^2] && (4.333)\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \underbrace{\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\sigma_{x_i}^2} - \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \mathbf{e}_i}_{\sigma_{x_i}^2} + \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\sigma_{x_i}^2} && (4.334)\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2. && (4.335)
\end{aligned}
$$

Hence, we end up with the optimization problem

$$
\mathbf{e}_{\mathrm{CWL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2 \right) \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\sigma_{x_i}^2} = 1. \tag{4.336}
$$

The optimization is started with the Lagrangian cost function, which is

$$
\mathcal{L}(\mathbf{e}_i) = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2 + \lambda \left(\frac{1}{\sigma_{x_i}^2} \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \mathbf{e}_i - 1 \right) + \lambda^* \left(\frac{1}{\sigma_{x_i}^2} \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}^* \mathbf{e}_i^* - 1 \right). \tag{4.337}
$$

The partial derivative of (4.337) w.r.t. $\mathbf{e}_i$ yields

$$
\frac{\partial \mathcal{L}(\mathbf{e}_i)}{\partial \mathbf{e}_i} = \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} + \lambda \frac{1}{\sigma_{x_i}^2} \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}. \tag{4.338}
$$

By setting (4.338) equal to zero, $\mathbf{e}_{\mathrm{CWL},i}^H$ can be derived as

$$
\mathbf{e}_{\mathrm{CWL},i}^H = -\lambda \frac{1}{\sigma_{x_i}^2} \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}. \tag{4.339}
$$

This result reinserted into the constraint in (4.336) leads to an expression for λ according to

$$\lambda = -\frac{(\sigma_{x_i}^2)^2}{\begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}. \tag{4.340}$$

Finally, reinserting (4.340) into (4.339) leads to the solution of the optimization problem in the form of

$$\mathbf{e}_{\mathrm{CWL},i}^H = \frac{\sigma_{x_i}^2}{\begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \tag{4.341}$$

$$= \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\underline{\mathbf{y}}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \mathbf{C}_{\underline{\mathbf{y}}x_i}} \mathbf{C}_{x_i\underline{\mathbf{y}}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}. \tag{4.342}$$

The full expression for $\hat{x}_{\mathrm{CWL},i}$ can be found by combining (4.74), (4.329) and (4.342), which yields

$$\hat{x}_{\mathrm{CWL},i} = E_{x_i}[x_i] + \mathbf{e}_{\mathrm{CWL},i}^H \big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big). \tag{4.343}$$

Using

$$\mathbf{E}_{\mathrm{CWL}} = \begin{bmatrix} \mathbf{e}_{\mathrm{CWL},1}^H \\ \mathbf{e}_{\mathrm{CWL},2}^H \\ \vdots \\ \mathbf{e}_{\mathrm{CWL},N_{\mathbf{x}}}^H \end{bmatrix} \in \mathbb{C}^{N_{\mathbf{x}} \times 2N_{\mathbf{y}}} \tag{4.344}$$

immediately leads us to the vector notation of the CWCU WLMMSE estimator

$$\hat{\mathbf{x}}_{\mathrm{CWL}} = E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{4.345}$$

According to (4.341), the CWCU WLMMSE estimator matrix $\mathbf{E}_{\mathrm{CWL}}$ can be derived from the augmented WLMMSE estimator matrix $\underline{\mathbf{E}}_{\mathrm{WL}} = \underline{\mathbf{C}}_{\mathbf{xy}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1}$ by

$$\mathbf{E}_{\mathrm{CWL}} = \begin{bmatrix} \mathbf{D} & \mathbf{0}^{N_{\mathbf{x}} \times N_{\mathbf{x}}} \end{bmatrix} \underline{\mathbf{E}}_{\mathrm{WL}}, \tag{4.346}$$

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

$$[\mathbf{D}]_{i,i} = \frac{\sigma_{x_i}^2}{\begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \underline{\mathbf{C}}_{\mathbf{yy}}^{-1} \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}. \tag{4.347}$$

Denoting $\widetilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & \mathbf{0}^{N_{\mathbf{x}} \times N_{\mathbf{x}}} \end{bmatrix}$, (4.348) reads as

$$\mathbf{E}_{\mathrm{CWL}} = \widetilde{\mathbf{D}} \underline{\mathbf{E}}_{\mathrm{WL}}. \tag{4.348}$$

We now derive some performance measures for the CWCU WLMMSE estimator. Starting with the mean of the error we have

$$E_{\mathbf{y},\mathbf{x}}[\hat{\mathbf{x}}_{\mathrm{CWL}} - \mathbf{x}] = E_{\mathbf{y},\mathbf{x}}\big[E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}\big(\mathbf{y} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \mathbf{x}\big] \tag{4.349}$$

$$= -E_{\mathbf{x}}[\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]] + \mathbf{E}_{\mathrm{CWL}} E_{\mathbf{y}}\big[\mathbf{y} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big] \tag{4.350}$$

$$= \mathbf{0}. \tag{4.351}$$

With that, the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{CWL}}$ follows as

$$\mathbf{C}_{\mathbf{ee},\mathrm{CWL}} = E_{\mathbf{y},\mathbf{x}}\big[\mathbf{e}\mathbf{e}^H\big] \tag{4.352}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \mathbf{x}\big)\big(E_{\mathbf{x}}[\mathbf{x}] + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \mathbf{x}\big)^H\Big] \tag{4.353}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\big) \times \big(-(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{E}_{\mathrm{CWL}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\big)^H\Big] \tag{4.354}$$

$$= \mathbf{C}_{\mathbf{xx}} - \mathbf{E}_{\mathrm{CWL}}\mathbf{C}_{\underline{\mathbf{y}}\mathbf{x}} - \mathbf{C}_{\mathbf{x}\underline{\mathbf{y}}}\mathbf{E}_{\mathrm{CWL}}^H + \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{E}_{\mathrm{CWL}}^H \tag{4.355}$$

$$= \mathbf{C}_{\mathbf{xx}} - \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yx}}\begin{bmatrix}\mathbf{I}^{n\times n}\\ \mathbf{0}^{n\times n}\end{bmatrix} - \begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}} & \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix}\underline{\mathbf{C}}_{\mathbf{xy}}\mathbf{E}_{\mathrm{CWL}}^H + \mathbf{E}_{\mathrm{CWL}}\underline{\mathbf{C}}_{\mathbf{yy}}\mathbf{E}_{\mathrm{CWL}}^H. \tag{4.356}$$

By utilizing $\underline{\mathbf{M}}$ as defined in (4.243), (4.356) simplifies to

$$\mathbf{C}_{\mathbf{ee},\mathrm{CWL}} = \mathbf{C}_{\mathbf{xx}} - \widetilde{\mathbf{D}}\underline{\mathbf{M}}\begin{bmatrix}\mathbf{I}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\\ \mathbf{0}^{N_{\mathbf{x}}\times N_{\mathbf{x}}}\end{bmatrix} - \begin{bmatrix}\mathbf{I}^{n\times n} & \mathbf{0}^{n\times n}\end{bmatrix}\underline{\mathbf{M}}\widetilde{\mathbf{D}}^H + \widetilde{\mathbf{D}}\underline{\mathbf{M}}\widetilde{\mathbf{D}}^H. \tag{4.357}$$

The BMSE values of the i^{th} estimate $\hat{x}_{\mathrm{CWL},i}$ corresponds to the i^{th} diagonal element of the error covariance matrix $\mathbf{C}_{\mathbf{ee},\mathrm{CWL}}$.

The findings of this section so far are summarized in case 1 of

Result 4.3 (CWCU WLMMSE Estimator for Real-Valued Parameter Vectors)

Let $\mathbf{y} \in \mathbb{C}^{N_\mathbf{y}}$. If $\mathbf{x} \in \mathbb{R}^{N_\mathbf{x}}$ is a real-valued parameter vector and

1. $\mathbf{x}$ and $\mathbf{y}_\mathbb{R} \in \mathbb{R}^{2N_\mathbf{y}}$ are jointly Gaussian, or
2. $\mathbf{x}$ and $\mathbf{y}$ are connected via the linear model in (2.6) 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
3. $\mathbf{x}$ and $\mathbf{y}$ are connected via the linear model in (2.6) and $\mathbf{x}$ has mean $E_\mathbf{x}[\mathbf{x}]$, mutually independent elements and covariance matrix $\mathbf{C}_{\mathbf{xx}} = \mathrm{diag}\{\sigma_{x_1}^2, \sigma_{x_2}^2, \cdots, \sigma_{x_{N_\mathbf{x}}}^2\}$ (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 $i = 1, 2, \cdots, N_\mathbf{x}$ is given by (4.345), where the estimator matrix $\mathbf{E}_{\mathrm{CWL}}$ is defined in (4.346) and (4.347). The mean of the error $\mathbf{e} = \hat{\mathbf{x}}_{\mathrm{CWL}} - \mathbf{x}$ (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 provided in (4.357) with $\underline{\mathbf{M}}$ defined in (4.243). The minimum BMSEs are $\mathrm{Bmse}(\hat{x}_{\mathrm{CWL},i}) = [\mathbf{M}_{\hat{\mathbf{x}}_{\mathrm{CWL}}}]_{i,i}$.

If none of the three cases is fulfilled, then in the linear model setup a widely linear CWCU estimator is available in form of the BWLUE for real-valued parameter vectors in Result 3.6, which not necessarily has to correspond to the CWCU WLMMSE estimator.

Case 2 and 3 in Result 4.3 origin from similar considerations as for case 2 and 3 in Result 4.1 and Result 4.2. The derivation of these cases can be found in Appendix Z.

4.3.4 Discussion of the CWCU WLMMSE Estimator for Real-Valued Parameters

Commonalities between the Three Cases in Result 4.3

For all three cases in Result 4.3 it holds that

$$E_{\mathbf{y}_\mathbb{R}|x_i}[\mathbf{y}_\mathbb{R}|x_i] = E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] + \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]). \tag{4.358}$$

The validity of this equation for the first case is clear and has already been utilized in (4.323). The proof for the other two cases is provided in Appendix AA. Multiplying

(4.358) with the real-to-complex transformation matrix $\mathbf{T}_{N_\mathbf{y}}$ defined in (2.3) from the left produces

$$E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] = E_\mathbf{y}[\underline{\mathbf{y}}] + \mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma_{x_i}^2}(x_i - E_{x_i}[x_i]), \tag{4.359}$$

which also holds for all three cases.

Similarly, it holds for all three cases that the conditional covariance matrix $\mathbf{C}_{\mathbf{y}_\mathbb{R}\mathbf{y}_\mathbb{R}|x_i}$ is given by

$$\mathbf{C}_{\mathbf{y}_\mathbb{R}\mathbf{y}_\mathbb{R}|x_i} = \mathbf{C}_{\mathbf{y}_\mathbb{R}\mathbf{y}_\mathbb{R}} - \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\mathbf{y}_\mathbb{R}}. \tag{4.360}$$

The proof is presented in Appendix AB. Multiplying (4.360) with $\mathbf{T}_{N_\mathbf{y}}$ from the left and with $\mathbf{T}_{N_\mathbf{y}}^H$ from the right results in

$$\underline{\mathbf{C}}_{\mathbf{yy}|x_i} = \underline{\mathbf{C}}_{\mathbf{yy}} - \mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\underline{\mathbf{y}}}. \tag{4.361}$$

Conditional Properties

The conditional properties of the ordinary BWLUE do not change compared to (4.251)–(4.254) since this classical estimator does not incorporate any statistics of the parameter vector. However, since the parameters are real-valued, it is beneficial to consider the BWLUE for real-valued parameter vectors from Result 3.6 instead. The derivation of its conditional properties is a straightforward extension of Appendix V when replacing $\underline{\mathbf{H}}$ with $\widetilde{\mathbf{H}}$ defined in (3.300). It leads to

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{BW},i}|x_i] = x_i, \tag{4.362}$$

$$b(\hat{x}_{\mathrm{BW},i}|x_i) = 0, \tag{4.363}$$

$$\mathrm{var}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathbf{u}_i^H\left(\widetilde{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\widetilde{\mathbf{H}}\right)^{-1}\mathbf{u}_i, \tag{4.364}$$

$$\mathrm{mse}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathbf{u}_i^H\left(\widetilde{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\widetilde{\mathbf{H}}\right)^{-1}\mathbf{u}_i. \tag{4.365}$$

Again, $\mathrm{var}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathrm{var}(\hat{x}_{\mathrm{BW},i})$ and $\mathrm{mse}(\hat{x}_{\mathrm{BW},i}|x_i) = \mathrm{mse}(\hat{x}_{\mathrm{BW},i})$ hold for the BWLUE.

Note that the WLMMSE estimator remains unaltered for real-valued parameter vectors. Only the utilized second order statistics change. However, the conditional properties provided in (4.255)–(4.258) do not hold any more since $\underline{\mathbf{D}}_i$ is singular for real-valued x_i. The modified conditional properties are

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] = [\mathbf{D}]_{i,i}^{-1}x_i + (1 - [\mathbf{D}]_{i,i}^{-1})E_{x_i}[x_i], \tag{4.366}$$

$$b(\hat{x}_{\mathrm{WL},i}|x_i) = \left([\mathbf{D}]_{i,i}^{-1} - 1\right)(x_i - E_{x_i}[x_i]), \tag{4.367}$$

$$\mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) = [\mathbf{D}]_{i,i}^{-1}\left(1 - [\mathbf{D}]_{i,i}^{-1}\right)\sigma_{x_i}^2, \tag{4.368}$$

$$\mathrm{mse}(\hat{x}_{\mathrm{WL},i}|x_i) = [\mathbf{D}]_{i,i}^{-1}\left(1 - [\mathbf{D}]_{i,i}^{-1}\right)\sigma_{x_i}^2 + \left(1 - [\mathbf{D}]_{i,i}^{-1}\right)^2|(x_i - E_{x_i}[x_i])|^2. \tag{4.369}$$

The according derivations can be found in Appendix AC. For deriving (4.366)–(4.369) we assumed that one of the three cases mentioned in Result 4.3 holds. Note the similarities between (4.366)–(4.369) and the conditional properties for the LMMSE estimator in (4.157)–(4.160). Interestingly, also the conditional properties of the CWCU WLMMSE estimator for real-valued parameters formally correspond to those of the CWCU LMMSE estimator and are given by

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] = x_i, \tag{4.370}$$
$$b(\hat{x}_{\mathrm{CWL},i}|x_i) = 0, \tag{4.371}$$
$$\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) = \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1), \tag{4.372}$$
$$\mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i) = \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). \tag{4.373}$$

The corresponding derivations can be found in Appendix AD. Due to these similarities, the expressions in (4.165)–(4.172) for the case of a zero mean parameter $E_{x_i}[x_i] = 0$ transformed to the case considered in this section directly follow as

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] = [\mathbf{D}]_{i,i} E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i], \tag{4.374}$$
$$\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) = [\mathbf{D}]_{i,i}^2 \mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i), \tag{4.375}$$
$$\mathrm{Bmse}(\hat{x}_{\mathrm{WL},i}) = \sigma_{x_i}^2 \left(1 - [\mathbf{D}]_{i,i}^{-1}\right), \tag{4.376}$$
$$\mathrm{Bmse}(\hat{x}_{\mathrm{CWL},i}) = \mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i) = \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1), \tag{4.377}$$
$$\mathrm{Bmse}(\hat{x}_{\mathrm{CWL},i}) = [\mathbf{D}]_{i,i} \mathrm{Bmse}(\hat{x}_{\mathrm{WL},i}), \tag{4.378}$$

and $\mathrm{Bmse}(\hat{x}_{\mathrm{CWL},i}) > \mathrm{Bmse}(\hat{x}_{\mathrm{WL},i})$. However, note the different definition of $[\mathbf{D}]_{i,i}$ utilized in Section 4.2.3.

Transformation Analysis

The CWCU WLMMSE estimator for real-valued parameters will in general not commute over affine transformations of the form

$$\boldsymbol{\alpha} = \mathbf{B}\mathbf{x} + \mathbf{c}. \tag{4.379}$$

The proof is a straightforward extension of the investigations in (4.288)–(4.297).

Again, an exception can be found. Let $\mathbf{c}$ be a zero vector and

$$\mathbf{B} = \mathrm{diag}\left\{\begin{bmatrix} b_1 & b_2 & \dots & b_{N_\mathbf{x}} \end{bmatrix}\right\} \in \mathbb{R}^{N_\mathbf{x} \times N_\mathbf{x}} \tag{4.380}$$

with non-zero diagonal elements. Then, $\mathbf{x} = \mathbf{B}^{-1}\boldsymbol{\alpha}$ holds. Furthermore, it holds that if one of the cases in Result 4.3 is fulfilled for $\mathbf{x}$, it is also fulfilled for $\boldsymbol{\alpha}$. Hence, Result 4.3

can be applied and leads to

$$\hat{\alpha}_{\mathrm{CWL},i} = E_{\alpha_i}[\alpha_i] + \frac{\sigma^2_{\alpha_i}}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{\alpha_i\mathbf{y}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\underline{\mathbf{C}}_{\mathbf{y}\alpha_i}\begin{bmatrix}1\\0\end{bmatrix}}\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{\alpha_i\mathbf{y}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) \tag{4.381}$$

$$= E_{\alpha_i}[\alpha_i] + \frac{\sigma^2_{\alpha_i}}{\mathbf{C}_{\alpha_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\mathbf{C}_{\underline{\mathbf{y}}\alpha_i}}\mathbf{C}_{\alpha_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big). \tag{4.382}$$

The statistics therein are given by

$$\alpha_i = b_i x_i + c_i, \tag{4.383}$$

$$E_{\alpha_i}[\alpha_i] = b_i E_{x_i}[x_i] + c_i, \tag{4.384}$$

$$\sigma^2_{\alpha_i} = b_i^2\sigma^2_{x_i}, \tag{4.385}$$

$$\mathbf{C}_{\underline{\mathbf{y}}\alpha_i} = E_{\mathbf{y},\alpha_i}\big[\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)(\alpha_i - E_{\alpha_i}[\alpha_i])^*\big], \tag{4.386}$$

$$= E_{\mathbf{y},x_i}\big[\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)(b_i(x_i - E_{x_i}[x_i]))^*\big], \tag{4.387}$$

$$= \mathbf{C}_{\underline{\mathbf{y}}x_i} b_i. \tag{4.388}$$

Inserting (4.384)–(4.388) into the CWCU WLMMSE estimator in (4.382) and utilizing the fact that b_i is non-zero results in

$$\hat{\alpha}_{\mathrm{CWL},i} = b_i E_{x_i}[x_i] + c_i + \frac{b_i^2\sigma^2_{x_i}}{b_i\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\mathbf{C}_{\underline{\mathbf{y}}x_i}b_i} b_i\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) \tag{4.389}$$

$$= b_i\left(E_{x_i}[x_i] + \frac{\sigma^2_{x_i}}{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\mathbf{C}_{\underline{\mathbf{y}}x_i}}\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\mathbf{yy}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\right) + c_i \tag{4.390}$$

$$= b_i\hat{x}_{\mathrm{CWL},i} + c_i. \tag{4.391}$$

This proves that the CWCU WLMMSE estimator for real-valued parameters commutes over affine transformations with real-valued diagonal transformation matrices.

4.3.5 PWCU WLMMSE Estimation

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

$$\mathbf{y} = \underbrace{\begin{bmatrix}\mathbf{H} & i\mathbf{H}\end{bmatrix}}_{\mathbf{H}' \in \mathbb{C}^{N_{\mathbf{y}} \times 2N_{\mathbf{x}}}}\underbrace{\begin{bmatrix}\mathbf{x}_{\mathrm{R}}\\ \mathbf{x}_{\mathrm{I}}\end{bmatrix}}_{\mathbf{x}_{\mathbb{R}} \in \mathbb{R}^{2N_{\mathbf{x}}}} + \mathbf{n}, \tag{4.392}$$

and estimate the real and imaginary parts of the parameter vector separately. With (4.392), the parameter vector is real-valued which enables us to use the CWCU WLMMSE estimator for real-valued parameter vectors given in Result 4.3. The estimated real and imaginary parts can then be combined to a complex estimator for the parameter vector $\mathbf{x}$. It is important 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 $i = 1, 2, \cdots, N_{\mathbf{x}}$. That is why this estimator will be denoted as part-wise conditionally unbiased (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 leads to weaker constraints as when conditioning on the complex parameters.

The derived estimators are compared with the classical BLUE and BWLUE as well as with the Bayesian LMMSE and WLMMSE estimators in the following example.

Example 4.7 (DC Level and Complex Exponential in Uncorrelated Gaussian Noise)

In this example we apply the CWCU WLMMSE estimator and the PWCU WLMMSE estimator to a particular signal parameter estimation problem, and compare their performance to that of 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 [59]. The signal model is $y[k] = x_1 + 1.5x_2 e^{j6k} + n[k]$ for $k = 0, 1, \cdots, 5$, which can easily be brought to the form of a linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$. We assume the noise vector $\mathbf{n}$ to be zero mean complex proper Gaussian with covariance matrix

$$\mathbf{C}_{\mathbf{nn}} = \mathrm{diag}\{0.1,\ 0.06,\ 0.3,\ 0.2,\ 0.15,\ 0.1\}. \tag{4.393}$$

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

$$\begin{aligned} \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}} &= \mathrm{diag}\{1,\ 0.6\}, && (4.394)\\ \mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}} &= k\,\mathrm{diag}\{1,\ 0.6\}, && (4.395)\\ \mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{I}} &= \mathbf{0}^{2\times 2}, && (4.396) \end{aligned}$$

where the scalar k in $\mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}}$ 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 almost purely real-valued), and test all the estimators listed in Table 4.2. 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.

Estimator	***Section***	***Equation/Result***
BLUE	3.1	(3.48)
LMMSE	4.1	(4.14)
CWCU LMMSE	4.2	Result 4.1
BWLUE	3.1	(3.63)
WLMMSE	4.1	(4.52)
CWCU WLMMSE for complex-valued parameter vectors	4.3.1	Result 4.2
CWCU WLMMSE for real-valued parameter vectors	4.3.3	Result 4.3
PWCU WLMMSE	4.3.5	Result 4.3

Table 4.2: Estimators used for the described estimation problem.

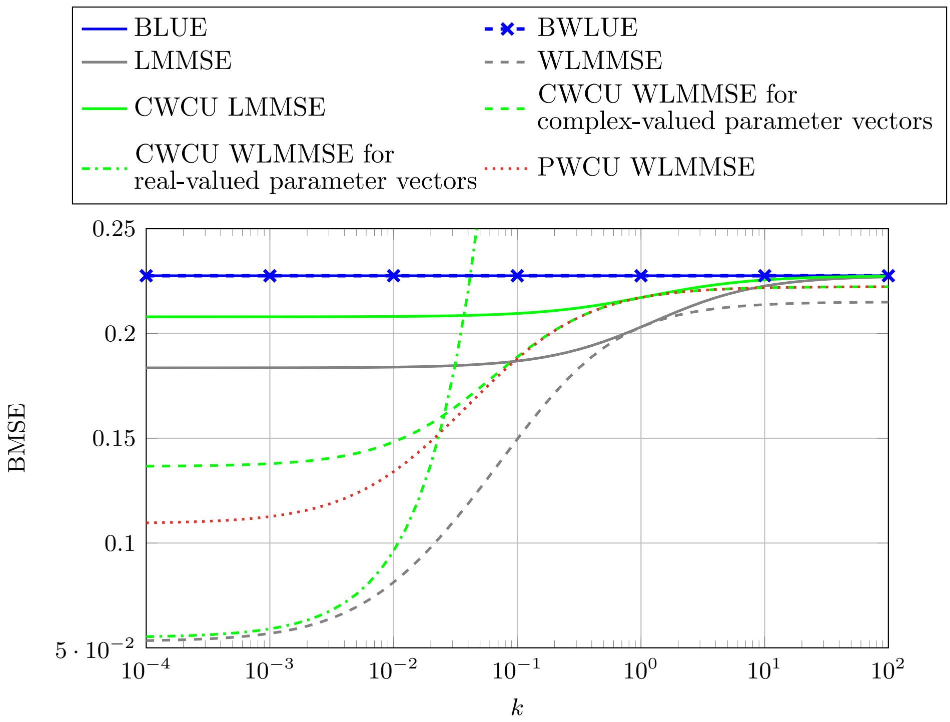

Figure 4.9: BMSE values plotted over the scaling factor k, which defines the variances of the imaginary parts. The variances of the real parts have been kept constant.

With this setup we can observe how the estimators perform for highly improper and also proper data within the scope of this example. Still, 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.

Figure 4.9 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. Clearly, for $k = 1$ (the proper case) they perform equal. This impressively shows that the LMMSE estimator is not able to exploit information about the improperness of $\mathbf{x}$. Further, the CWCU WLMMSE estimator derived in this work also significantly outperforms the LMMSE estimator for small values of k, and it is also better 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 when $k \to -\infty$. Consequently, once we know from the application that the parameter vector is real one 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 complex-valued parameters for small k.

We already noted that for $k = 1$ (the proper case), the LMMSE and the WLMMSE estimators perform equal. The same holds 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 $\mathbf{C}_{\mathbf{x}_\mathrm{I}\mathbf{x}_\mathrm{I}}$ become less important. What's left is the prior knowledge about $\mathbf{C}_{\mathbf{x}_\mathrm{R}\mathbf{x}_\mathrm{R}}$. 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 for complex-valued parameters 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 estimators significantly outperform their 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.

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

Table 4.3: Linear and widely linear estimators and their constraints

4.4 Estimator Comparison

In Section 3.1, the classical BLUE was derived by minimizing the estimators variance subject to the unbiased constraint $E_{\mathbf{y}}[\hat{x}_i] = x_i$. This unbiased constraint led to estimator matrices fulfilling $\mathbf{EH} = \mathbf{I}$. In Appendix J, we showed that the BLUE can also be derived in a Bayesian framework by minimizing the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ subject to the same unbiased constraint. Similar arguments also hold for the BWLUE. Hence, every estimator regarded in Chapter 4 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

$$\hat{\mathbf{x}} = \mathbf{F}\mathbf{y} + \mathbf{G}\mathbf{y}^* = \begin{bmatrix}\mathbf{F} & \mathbf{G}\end{bmatrix}\underline{\mathbf{y}}. \tag{4.397}$$

Table 4.3 lists the main estimators regarded in Chapter 4 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 perform better in a BMSE sense than the BLUE and the BWLUE, respectively. Yet, they are conditionally biased, leading to effects demonstrated in Example 4.1. The CWCU estimators derived in this paper circumvent this property. Thus, in contrast to the BLUE and the BWLUE, the proposed estimators are generally able to incorporate prior knowledge about the statistics of the parameter vector. This can lead to a significant performance gain over classical estimators.

5 Knowledge-Aided Concepts in Adaptive Filtering

Adaptive filters are utilized in many practical applications. Most of these applications can be divided into the four groups [71,72]

- System identification
- Noise cancellation
- Inverse system identification
- Prediction.

In this chapter well-known complex-valued adaptive filters such as the LMS and RLS algorithms are shortly recapitulated. Subsequently, we extend these adaptive filters by incorporating additional sources of knowledge that are available in many practical scenarios.

Firstly, we consider the case when it is known that the optimum filter coefficients are real-valued, whereas the input and desired signal are complex-valued. A practical example where this situation can arise is given by the already mentioned problem of transmit leakage in modern wireless transceivers. In [16] the leakage signal is extracted and cancelled by using a so-called auxiliary receiver in parallel to the main receiver. As a result, typically a fractional delay between the two receivers appears. This fractional delay is estimated and compensated for with the help of adaptive filters. In this scenario, the input and desired signal are complex-valued while the optimum filter coefficients are real-valued. Standard complex-valued adaptive filters cannot incorporate the additional model knowledge and produce complex-valued filter coefficients. In this thesis, extensions of the LMS and RLS algorithms are developed that produce real-valued filter coefficients. These optimal filters are compared with state-of-the-art filters as well as with trivial filters that incorporate the additional model knowledge in an intuitive way. For the LMS-based algorithms it turns out that the utilized intuitive filter corresponds to the optimal approach derived in this thesis. In case of the RLS-based algorithms, however, the derived optimal algorithm outperforms an intuitive filter significantly as will be demonstrated in simulation examples.

Subsequently, the system identification application of adaptive filters is considered. There, prior statistical knowledge about the impulse response to be estimated is often available. Hence, adaptive filters are investigated that are able to incorporate this prior knowledge. An adaptive filter that is able to incorporate prior knowledge is the sequential LMMSE estimator in a filtering setup. This algorithm is related to the RLS algorithm. A novel version of the LMS algorithm that also allows to incorporate the first and second order statistical moments about the impulse response to be estimated has been derived in the context of this doctoral thesis work. It is shown that the derived LMS-based algorithm incorporating prior knowledge features a reduced convergence time in the mean compared to the standard LMS algorithm.

5.1 State-of-the-Art

Figure 5.1 schematically depicts the task of adaptive filter based system identification, where $\mathbf{h} \in \mathbb{C}^{N_\mathbf{h}}$ is the unknown impulse response of a LTI system, $x_k \in \mathbb{C}$ is the known input of the system at time instance k, $y_k \in \mathbb{C}$ is the measured output of the system at time instance k, and $n_k \in \mathbb{C}$ is an unknown noise sample. The samples y_k are usually referred to as desired signal. The adaptive filter with time-dependent impulse response $\mathbf{w}_k \in \mathbb{C}^{N_\mathbf{w}}$ is fed with the same input samples x_k and produces $\hat{y}_k \in \mathbb{C}$ as output samples. The error $e_k = y_k - \hat{y}_k$ is used to adapt the filter coefficients $\mathbf{w}_k$.

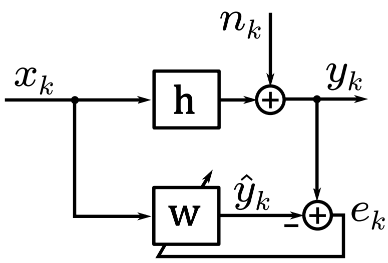

Figure 5.1: System identification example with an adaptive filter.

The following considerations are independent of the particular adaptive filter applications and are not restricted to the system identification problem.

There exist several notations of deriving the output of a complex-valued filter [72]. The form

$$\hat{y}_k = \mathbf{w}^T \mathbf{x}_k \tag{5.1}$$

is preferred in this work, where $\mathbf{x}_k \in \mathbb{C}^{N_\mathbf{w}}$ contains the latest $N_\mathbf{w}$ samples of the input signal x_k as

$$\mathbf{x}_k = \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-N_\mathbf{w}+1} \end{bmatrix}. \tag{5.2}$$

Other frequently used notations are $\hat{y}_k = \mathbf{w}^H \mathbf{x}_k$ and $\hat{y}_k = \mathbf{w}^T \mathbf{x}_k^*$. Note that the obtained adaptive filter update equations depend on the chosen notation.

Wiener-Hopf Solution

For complex-valued input signals and complex-valued impulse responses, the famous Wiener-Hopf solution for the particular choice in (5.1) is given by

$$\mathbf{w} = \left(\mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{r}_{\mathbf{x}y}\right)^*, \tag{5.3}$$

where the auto-correlation matrix $\mathbf{R}_{\mathbf{xx}}$ is defined as

$$\mathbf{R}_{\mathbf{xx}} = E_{\mathbf{x}_k}[\mathbf{x}_k \mathbf{x}_k^H] \in \mathbb{C}^{N_\mathbf{w} \times N_\mathbf{w}} \tag{5.4}$$

and where the cross-correlation vector between $\mathbf{x}_k$ and y_k is defined as

$$\mathbf{r}_{\mathbf{x}y} = E_{\mathbf{x}_k, y_k}[\mathbf{x}_k y_k^*] \in \mathbb{C}^{N_\mathbf{w}}. \tag{5.5}$$

Applying the Wiener-Hopf solution requires the knowledge of the statistics in (5.4) and (5.5). This can be avoided by utilizing adaptive filters. The most commonly used adaptive filters can be considered to be the LMS [73] and the RLS algorithms [71], whose derivations are now repeated for the complex-valued case.

LMS Algorithm

We first take a look at the LMS algorithm for real-valued signals and systems. This algorithm adaptively minimizes a cost function J utilizing the *steepest descent* method of the form[5]

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu \left(\frac{\partial J}{\partial \mathbf{w}_{k-1}} \right)^T \tag{5.6}$$

to derive the next filter coefficient estimate $\mathbf{w}_k$ based on the old estimate $\mathbf{w}_{k-1}$. The step-size is denoted by $\mu \in \mathbb{R}$ in (5.6). The transpose operator in (5.6) is necessary since the partial derivative produces a row vector per definition.

We now turn to complex-valued signals and filters. Let $\mathbf{w}_{\mathrm{R},k}$ and $\mathbf{w}_{\mathrm{I},k}$ denote the real and imaginary part of $\mathbf{w}_k$, respectively. Then, we seek for two real-valued LMS algorithms of the form

$$\mathbf{w}_{\mathrm{R},k} = \mathbf{w}_{\mathrm{R},k-1} - \mu \left(\frac{\partial J}{\partial \mathbf{w}_{\mathrm{R},k-1}} \right)^T, \qquad \mathbf{w}_{\mathrm{I},k} = \mathbf{w}_{\mathrm{I},k-1} - \mu \left(\frac{\partial J}{\partial \mathbf{w}_{\mathrm{I},k-1}} \right)^T. \tag{5.7}$$

[5] Another usual notation of the gradient term is $\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}\big|_{\mathbf{w}=\mathbf{w}_{k-1}}$

Recombining these two adaptive filters to a single complex-valued adaptive filter produces

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu\left(\frac{\partial J}{\partial \mathbf{w}_{\mathrm{R},k-1}} + j\frac{\partial J}{\partial \mathbf{w}_{\mathrm{I},k-1}}\right)^T. \tag{5.8}$$

Interestingly, by comparing the derivatives in (5.8) with (2.37) and (2.47) reveals that they correspond to the Wirtinger derivative of J w.r.t. $\mathbf{w}_{k-1}^*$ scaled by a factor of 2. Hence, for complex-valued adaptive filters (5.6) needs to be adapted according to

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu\left(\frac{\partial J}{\partial \mathbf{w}_{k-1}*}\right)^T, \tag{5.9}$$

where the scaling factor is moved into the step-size μ. The LMS algorithm utilizes the absolute squared instantaneous error as cost function, which is

$$\begin{aligned}
J &= |e_k|^2 && (5.10)\\
&= e_k e_k^* && (5.11)\\
&= (y_k - \hat{y}_k)(y_k - \hat{y}_k)^* && (5.12)\\
&= \left(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\right)\left(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\right)^* && (5.13)\\
&= y_k y_k^* - \mathbf{w}_{k-1}^T\mathbf{x}_k y_k^* - y_k\mathbf{w}_{k-1}^H\mathbf{x}_k^* + \mathbf{w}_{k-1}^T\mathbf{x}_k\mathbf{w}_{k-1}^H\mathbf{x}_k^* && (5.14)\\
&= y_k y_k^* - \mathbf{w}_{k-1}^T\mathbf{x}_k y_k^* - \mathbf{w}_{k-1}^H\mathbf{x}_k^* y_k + \mathbf{w}_{k-1}^H\mathbf{x}_k^*\mathbf{x}_k^T\mathbf{w}_{k-1}. && (5.15)
\end{aligned}$$

The Wirtinger derivative of (5.15) w.r.t. $\mathbf{w}_{k-1}^*$ is given by

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{w}_{k-1}^*} &= -\, y_k\mathbf{x}_k^H + \mathbf{w}_{k-1}^T\mathbf{x}_k\mathbf{x}_k^H && (5.16)\\
&= -\, y_k\mathbf{x}_k^H + \hat{y}_k\mathbf{x}_k^H && (5.17)\\
&= -\, e_k\mathbf{x}_k^H. && (5.18)
\end{aligned}$$

Inserting (5.18) into (5.9) yields the final result for the LMS algorithm

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu e_k\mathbf{x}_k^*. \tag{5.19}$$

As initialization, $\mathbf{w}_0$ is in many cases chosen to be the zero vector. One can show that for convergence the step-size μ must be chosen to be [72]

$$0 < \mu < \frac{2}{\lambda_{\max}}, \tag{5.20}$$

where $\lambda_{\max}$ is the largest eigenvalue of the auto-correlation matrix $\mathbf{R}_{\mathbf{xx}} = E_{\mathbf{x}_k}[\mathbf{x}_k\mathbf{x}_k^H] \in \mathbb{C}^{N_\mathbf{w}\times N_\mathbf{w}}$ [72]. Furthermore, one can show that the LMS algorithm in the mean converges to the Wiener-Hopf solution in (5.3).

A more practical criterion for the choice of μ is given by [72]

$$0 < \mu < \frac{2}{N_\mathbf{w}E[|x_k|^2]} = \frac{2}{E[\mathbf{x}_k^H\mathbf{x}_k]}. \tag{5.21}$$

Normalized LMS Algorithm

Approximating $E[\mathbf{x}_k^H \mathbf{x}_k]$ in(5.21) by the instantaneous value $\|\mathbf{x}_k\|_2^2$ motivates the so called normalized LMS algorithm with its update equation

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \frac{\mu_n}{\epsilon + \|\mathbf{x}_k\|_2^2} e_k \mathbf{x}_k^*. \tag{5.22}$$

In practice, μ_n is typically chosen in the range between 0 and 1. $\epsilon \in \mathbb{R}$ in (5.22) is a small positive-valued constant to overcome possible instabilities when $\|\mathbf{x}_k\|_2^2$ is very small. This algorithm is also sometimes referred to as ϵ-normalized least mean squares (NLMS) to emphasize the incorporation of ϵ [71].

Finally, we emphasize that the LMS as well as the NLMS algorithms are often implemented with a variable step-size μ_k that depends on the time index k.

RLS Algorithm

We now recapitulate the derivation of the RLS algorithm for the complex-valued case, for which the following cost function is utilized:

$$J = \sum_{i=0}^{k} \lambda^{k-i} |e_i|^2. \tag{5.23}$$

This cost function can be rewritten as

$$\begin{aligned} J &= \sum_{i=0}^{k} \lambda^{k-i} e_i e_i^* && (5.24) \\ &= \sum_{i=0}^{k} \lambda^{k-i} (y_i - \hat{y}_i)(y_i - \hat{y}_i)^*. && (5.25) \end{aligned}$$

In (5.25) we use $\hat{y}_i = \mathbf{w}_k^T \mathbf{x}_i$ such that the weighted sum of squared errors is calculated based on the current filter coefficients $\mathbf{w}_k$. Hence, we write

$$\begin{aligned} J &= \sum_{i=0}^{k} \lambda^{k-i} \big(y_i - \mathbf{w}_k^T \mathbf{x}_i\big)\big(y_i - \mathbf{w}_k^T \mathbf{x}_i\big)^* && (5.26) \\ &= \sum_{i=0}^{k} \lambda^{k-i} \big(y_i y_i^* - \mathbf{w}_k^T \mathbf{x}_i y_i^* - y_i \mathbf{w}_k^H \mathbf{x}_i^* + \mathbf{w}_k^T \mathbf{x}_i \mathbf{w}_k^H \mathbf{x}_i^*\big) && (5.27) \\ &= \sum_{i=0}^{k} \lambda^{k-i} \big(y_i y_i^* - \mathbf{w}_k^T \mathbf{x}_i y_i^* - \mathbf{w}_k^H \mathbf{x}_i^* y_i + \mathbf{w}_k^H \mathbf{x}_i^* \mathbf{x}_i^T \mathbf{w}_k\big). && (5.28) \end{aligned}$$

The Wirtinger derivative of (5.28) w.r.t. $\mathbf{w}_k^*$ follows as

$$\frac{\partial J}{\partial \mathbf{w}_k^*} = \sum_{i=0}^{k} \lambda^{k-i}\big(-y_i \mathbf{x}_i^H + \mathbf{w}_k^T \mathbf{x}_i \mathbf{x}_i^H\big) \tag{5.29}$$

$$= \sum_{i=0}^{k} \lambda^{k-i}\big(-y_i + \mathbf{w}_k^T \mathbf{x}_i\big)\mathbf{x}_i^H \tag{5.30}$$

$$= \mathbf{w}_k^T \sum_{i=0}^{k} \lambda^{k-i} \mathbf{x}_i \mathbf{x}_i^H - \sum_{i=0}^{k} \lambda^{k-i} y_i \mathbf{x}_i^H. \tag{5.31}$$

We now introduce the following definitions

$$\mathbf{R}_k = \sum_{i=0}^{k} \lambda^{k-i} \mathbf{x}_i \mathbf{x}_i^H \tag{5.32}$$

and

$$\mathbf{r}_k^H = \sum_{i=0}^{k} \lambda^{k-i} y_i \mathbf{x}_i^H. \tag{5.33}$$

Then, setting (5.31) equal to zero results in

$$\mathbf{w}_k^T = \mathbf{r}_k^H \mathbf{R}_k^{-1}. \tag{5.34}$$

The next step is to find expressions for $\mathbf{r}_k$ and $\mathbf{R}_k$ in dependence of $\mathbf{r}_{k-1}$ and $\mathbf{R}_{k-1}$ in order to allow for recursive evaluations. We begin with the first one and rewrite (5.33) according to

$$\mathbf{r}_k^H = y_k \mathbf{x}_k^H + \sum_{i=0}^{k-1} \lambda^{k-i} y_i \mathbf{x}_i^H \tag{5.35}$$

$$= y_k \mathbf{x}_k^H + \lambda \sum_{i=0}^{k-1} \lambda^{k-1-i} y_i \mathbf{x}_i^H \tag{5.36}$$

$$= y_k \mathbf{x}_k^H + \lambda \mathbf{r}_{k-1}^H. \tag{5.37}$$

For (5.32) it holds that

$$\mathbf{R}_k = \mathbf{x}_k \mathbf{x}_k^H + \sum_{i=0}^{k-1} \lambda^{k-i} \mathbf{x}_i \mathbf{x}_i^H \tag{5.38}$$

$$= \mathbf{x}_k \mathbf{x}_k^H + \lambda \sum_{i=0}^{k-1} \lambda^{k-1-i} \mathbf{x}_i \mathbf{x}_i^H \tag{5.39}$$

$$= \mathbf{x}_k \mathbf{x}_k^H + \lambda \mathbf{R}_{k-1}. \tag{5.40}$$

In principle, a recursive algorithm can already be obtained by combining (5.37) and (5.40) with (5.34). However, the recursion would contain a matrix inversion, which is avoided by the following reformulations.

We apply Woodbury's matrix inversion lemma [54] on (5.40) to yield

$$(\mathbf{R}_k)^{-1} = (\lambda\mathbf{R}_{k-1})^{-1} - (\lambda\mathbf{R}_{k-1})^{-1}\mathbf{x}_k\Big(1 + \mathbf{x}_k^H(\lambda\mathbf{R}_{k-1})^{-1}\mathbf{x}_k\Big)^{-1}\mathbf{x}_k^H(\lambda\mathbf{R}_{k-1})^{-1}. \quad (5.41)$$

Denoting $(\mathbf{R}_k)^{-1}$ as $\mathbf{P}_k$ produces

$$\mathbf{P}_k = \lambda^{-1}\mathbf{P}_{k-1} - \lambda^{-1}\mathbf{P}_{k-1}\mathbf{x}_k\big(1 + \lambda^{-1}\mathbf{x}_k^H\mathbf{P}_{k-1}\mathbf{x}_k\big)^{-1}\mathbf{x}_k^H\lambda^{-1}\mathbf{P}_{k-1}. \quad (5.42)$$

We now introduce the *gain vector* as

$$\mathbf{g}_k = \lambda^{-1}\mathbf{P}_{k-1}\mathbf{x}_k\big(1 + \lambda^{-1}\mathbf{x}_k^H\mathbf{P}_{k-1}\mathbf{x}_k\big)^{-1}, \quad (5.43)$$

such that (5.42) follows as

$$\begin{aligned}\mathbf{P}_k &= \lambda^{-1}\mathbf{P}_{k-1} - \mathbf{g}_k\mathbf{x}_k^H\lambda^{-1}\mathbf{P}_{k-1} && (5.44)\\ &= \lambda^{-1}\big(\mathbf{P}_{k-1} - \mathbf{g}_k\mathbf{x}_k^H\mathbf{P}_{k-1}\big). && (5.45)\end{aligned}$$

Since $\mathbf{P}_k = (\mathbf{R}_k)^{-1}$ is a Hermitian matrix according to (5.32), (5.45) can be expressed as

$$\mathbf{P}_k = \lambda^{-1}\big(\mathbf{P}_{k-1} - \mathbf{P}_{k-1}\mathbf{x}_k\mathbf{g}_k^H\big). \quad (5.46)$$

Furthermore, a reformulation of (5.43) yields

$$\begin{aligned}\mathbf{g}_k\big(1 + \lambda^{-1}\mathbf{x}_k^H\mathbf{P}_{k-1}\mathbf{x}_k\big) &= \lambda^{-1}\mathbf{P}_{k-1}\mathbf{x}_k && (5.47)\\ \mathbf{g}_k + \mathbf{g}_k\lambda^{-1}\mathbf{x}_k^H\mathbf{P}_{k-1}\mathbf{x}_k &= \lambda^{-1}\mathbf{P}_{k-1}\mathbf{x}_k && (5.48)\\ \mathbf{g}_k &= \lambda^{-1}\mathbf{P}_{k-1}\mathbf{x}_k - \mathbf{g}_k\lambda^{-1}\mathbf{x}_k^H\mathbf{P}_{k-1}\mathbf{x}_k && (5.49)\\ \mathbf{g}_k &= \lambda^{-1}\big(\mathbf{P}_{k-1} - \mathbf{g}_k\mathbf{x}_k^H\mathbf{P}_{k-1}\big)\mathbf{x}_k. && (5.50)\end{aligned}$$

The right hand side of (5.50) can be identified as $\mathbf{P}_k\mathbf{x}_k$ according to (5.45) such that

$$\mathbf{g}_k = \mathbf{P}_k\mathbf{x}_k. \quad (5.51)$$

We are now able to find an expression for $\mathbf{w}_k$ in (5.34)

$$\begin{aligned}\mathbf{w}_k^T &= \mathbf{r}_k^H\mathbf{R}_k^{-1} && (5.52)\\ &= \mathbf{r}_k^H\mathbf{P}_k. && (5.53)\end{aligned}$$

Incorporating (5.37) into (5.53) produces

$$\begin{aligned}\mathbf{w}_k^T &= y_k\mathbf{x}_k^H\mathbf{P}_k + \lambda\mathbf{r}_{k-1}^H\mathbf{P}_k && (5.54)\\ &= \lambda\mathbf{r}_{k-1}^H\mathbf{P}_k + y_k\mathbf{g}_k^H, && (5.55)\end{aligned}$$

where we utilized (5.51) and the fact that $\mathbf{P}_k = (\mathbf{R}_k)^{-1}$ is a Hermitian matrix according to (5.32). Combining the recursive definition of $\mathbf{P}_k$ in (5.46) with (5.55) allows for

$$\begin{aligned}\mathbf{w}_k^T &= \mathbf{r}_{k-1}^H\big(\mathbf{P}_{k-1} - \mathbf{P}_{k-1}\mathbf{x}_k\mathbf{g}_k^H\big) + y_k\mathbf{g}_k^H && (5.56)\\ &= \mathbf{w}_{k-1}^T - \mathbf{w}_{k-1}^T\mathbf{x}_k\mathbf{g}_k^H + y_k\mathbf{g}_k^H && (5.57)\\ &= \mathbf{w}_{k-1}^T + \big(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\big)\mathbf{g}_k^H && (5.58)\\ &= \mathbf{w}_{k-1}^T + e_k\mathbf{g}_k^H, && (5.59)\end{aligned}$$

with the a-priori error

$$e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k. \tag{5.60}$$

This is the final form of the recursion of the RLS algorithm. The filter coefficients $\mathbf{w}_0$ are usually initialized as a zero vector. The forgetting factor λ is chosen between 0.98 and 1 in most cases. $\mathbf{P}_0$ will be initialized as a scaled identity matrix $\delta\mathbf{I}$ with a large value of δ.

Finally, the RLS algorithm is summarized as

Initialization:
  Choose λ and δ;
  $\mathbf{P}_0 = \delta\mathbf{I}^{N_\mathbf{w} \times N_\mathbf{w}}$;
  $\mathbf{w}_0 = \mathbf{0}^{N_\mathbf{w} \times 1}$;
for $k = 1, 2, \ldots$ **do**
  Update $\mathbf{x}_k$ according to (5.2);
  Derive e_k according to (5.60);
  Determine $\mathbf{g}_k$ according to (5.43);
  Derive $\mathbf{P}_k$ according to (5.45);
  Evaluate the new filter coefficients $\mathbf{w}_k$ according to (5.59);
end

The novel adaptive filters introduced in the next sections will be compared in performance with the LMS and RLS algorithms described in this section.

5.2 Adaptive Filters for Real-Valued Filter Coefficients in Complex-Valued Environments

In Section 3.4, we investigated estimators that are designed for estimating real-valued parameter vectors while the measurement matrix and the measurement noise are both complex-valued. These estimators offered a significant performance gain compared to competing well-known and intuitive estimation methods. In this section, similar extensions are investigated for adaptive filters such as the LMS and the RLS algorithms for scenarios where it is known that the filter coefficients should be real valued, while the input signal x_k and the desired signal y_k are complex-valued. Examples and a potential application are given at the end of the section. At first, the LMS algorithm for real-valued filter coefficients is derived. After that, a similar extension is investigated for the RLS algorithm.

Note that the LMS and the RLS algorithms discussed in Section 5.1 do not incorporate the additional knowledge that the filter coefficients shall be real-valued, while the LMS and the RLS algorithms derived in the following do so. To not confuse the reader with varying notations, we refer to the algorithms described in Section 5.1 as the ordinary LMS algorithm and the ordinary RLS algorithm.

The optimal Wiener filter for the described scenario is not given by (5.3) any more. Instead, the optimal Wiener filter $\mathbf{w}$ is derived in the following. The cost function is given by

$$
\begin{aligned}
J &= E\Big[|e_k|^2\Big] && (5.61)\\
&= E[e_k e_k^*] && (5.62)\\
&= E[(y_k - \hat{y}_k)(y_k - \hat{y}_k)^*] && (5.63)\\
&= E\Big[\big(y_k - \mathbf{w}^T\mathbf{x}_k\big)\big(y_k - \mathbf{w}^T\mathbf{x}_k\big)^*\Big]. && (5.64)
\end{aligned}
$$

Incorporating the fact that $\mathbf{w}$ shall be real-valued allows for

$$
\begin{aligned}
J &= E\big[y_k y_k^* - \mathbf{w}^T\mathbf{x}_k y_k^* - y_k\mathbf{x}_k^H\mathbf{w} + \mathbf{w}^T\mathbf{x}_k\mathbf{x}_k^H\mathbf{w}\big] && (5.65)\\
&= E[y_k y_k^*] - \mathbf{w}^T\mathbf{r}_{\mathbf{x}y} - \mathbf{r}_{\mathbf{x}y}^H\mathbf{w} + \mathbf{w}^T\mathbf{R}_{\mathbf{xx}}\mathbf{w} && (5.66)
\end{aligned}
$$

The partial derivative of (5.66) w.r.t. $\mathbf{w}$ is

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{w}} &= -\mathbf{r}_{\mathbf{x}y}^T - \mathbf{r}_{\mathbf{x}y}^H + \mathbf{w}^T\big(\mathbf{R}_{\mathbf{xx}} + \mathbf{R}_{\mathbf{xx}}^T\big) && (5.67)\\
&= -\mathbf{r}_{\mathbf{x}y}^T - \mathbf{r}_{\mathbf{x}y}^H + \mathbf{w}^T(\mathbf{R}_{\mathbf{xx}} + \mathbf{R}_{\mathbf{xx}}^*) && (5.68)\\
&= -2\mathrm{Re}\big\{\mathbf{r}_{\mathbf{x}y}^T\big\} + 2\mathbf{w}^T\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}. && (5.69)
\end{aligned}
$$

Setting (5.69) equal to zero allows to identify the optimal Wiener filter for the case of real-valued filter coefficients as

$$
\mathbf{w} = (\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\})^{-1}\mathrm{Re}\{\mathbf{r}_{\mathbf{x}y}\}, \tag{5.70}
$$

For the system identification task in Figure 5.1, it holds that $\mathbf{r}_{\mathbf{x}y} = \mathbf{R}_{\mathbf{xx}}\mathbf{h}$. Let the true impulse response $\mathbf{h}$ be real-valued. Then, modifying (5.70) shows that

$$
\begin{aligned}
\mathbf{w} &= (\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\})^{-1}\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\mathbf{h}\} && (5.71)\\
&= (\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\})^{-1}\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}\mathbf{h} && (5.72)\\
&= \mathbf{h}. && (5.73)
\end{aligned}
$$

5.2.1 LMS Algorithm for Real-Valued Filter Coefficients

Consider the update equation in (5.6). This update equation, although it is usually applied for pure real-valued models, remains valid for the described scenario since $\mathbf{w}_k$ is real-valued. We now modify the partial derivative in this update equation in an optimal way such that only real-valued filter coefficients $\mathbf{w}_k$ are obtained. The LMS cost function is defined to be the instantaneous absolute squared error

$$
\begin{aligned}
J &= |e_k|^2 && (5.74)\\
&= e_k e_k^* && (5.75)\\
&= (y_k - \hat{y}_k)(y_k - \hat{y}_k)^* && (5.76)\\
&= \big(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\big)\big(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\big)^*. && (5.77)
\end{aligned}
$$

Assuming that $\mathbf{w}_{k-1}$ is real-valued allows to further modify (5.77) as

$$J = y_k y_k^* - \mathbf{w}_{k-1}^T \mathbf{x}_k y_k^* - \mathbf{w}_{k-1}^T \mathbf{x}_k^* y_k + \mathbf{w}_{k-1}^T \mathbf{x}_k \mathbf{x}_k^H \mathbf{w}_{k-1}. \tag{5.78}$$

For the update equation in (5.6), the partial derivative of (5.78) is required, which can be derived as

$$\frac{\partial J}{\partial \mathbf{w}_{k-1}} = -y_k^* \mathbf{x}_k^T - y_k \mathbf{x}_k^H + \mathbf{w}_{k-1}^T \big(\mathbf{x}_k \mathbf{x}_k^H + \mathbf{x}_k^* \mathbf{x}_k^T\big). \tag{5.79}$$

Note that no Wirtinger calculus for determining the partial derivative is necessary since $\mathbf{w}_{k-1}$ is real-valued. Inserting (5.79) into (5.6) yields

$$\begin{aligned}
\mathbf{w}_k &= \mathbf{w}_{k-1} - \mu\big(-\mathbf{x}_k y_k^* - \mathbf{x}_k^* y_k + \big(\mathbf{x}_k \mathbf{x}_k^H + \mathbf{x}_k^* \mathbf{x}_k^T\big)\mathbf{w}_{k-1}\big) && (5.80)\\
&= \mathbf{w}_{k-1} + \mu\big(\mathbf{x}_k y_k^* + \mathbf{x}_k^* y_k - \mathbf{x}_k \mathbf{x}_k^H \mathbf{w}_{k-1} - \mathbf{x}_k^* \mathbf{x}_k^T \mathbf{w}_{k-1}\big) && (5.81)\\
&= \mathbf{w}_{k-1} + \mu(\mathbf{x}_k y_k^* + \mathbf{x}_k^* y_k - \mathbf{x}_k \hat{y}_k^* - \mathbf{x}_k^* \hat{y}_k) && (5.82)\\
&= \mathbf{w}_{k-1} + \mu(\mathbf{x}_k e_k^* + \mathbf{x}_k^* e_k) && (5.83)\\
&= \mathbf{w}_{k-1} + 2\mu \mathrm{Re}\{e_k \mathbf{x}_k^*\}. && (5.84)
\end{aligned}$$

The additional factor of 2 will be moved into the step-size μ such that the final update equation is given as

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu \mathrm{Re}\{e_k \mathbf{x}_k^*\}. \tag{5.85}$$

Inspecting this result reveals that it is similar to the ordinary LMS algorithm in (5.19), except that only the real part of $e_k \mathbf{x}_k^*$ is used to update the filter coefficients. An intuitive approach of incorporating the fact that the filter coefficients should be real-valued might have led to the same result. However, the derivation here shows that this intuitive approach turns out to be optimal in terms of the instantaneous error cost function in (5.10).

Although (5.85) is similar to the ordinary LMS algorithm, the convergence solution and the stability analysis reveal some significant differences.

Convergence Solution

Let's assume μ is chosen such that the algorithm converges in the mean. Bounds for μ such that convergence is reached in the mean are derived in the next subsection. Here, we ask ourself about which $\mathbf{w}$ the algorithm will converge in the mean.

When the algorithm has reached its "convergence in the mean" state (for large enough k), it must hold that

$$E[\mathrm{Re}\{e_k \mathbf{x}_k^*\}] = \mathbf{0}. \tag{5.86}$$

Reformulating (5.86) results in

$$E[\mathrm{Re}\{e_k\mathbf{x}_k^*\}] = E\left[\frac{1}{2}(e_k\mathbf{x}_k^* + e_k^*\mathbf{x}_k)\right] \tag{5.87}$$

$$= E\left[\frac{1}{2}(y_k\mathbf{x}_k^* - \hat{y}_k\mathbf{x}_k^* + y_k^*\mathbf{x}_k - \hat{y}_k^*\mathbf{x}_k)\right] \tag{5.88}$$

$$= E\left[\frac{1}{2}\big(y_k\mathbf{x}_k^* - (\mathbf{w}_{k-1}^T\mathbf{x}_k)\mathbf{x}_k^* + y_k^*\mathbf{x}_k - (\mathbf{w}_{k-1}^T\mathbf{x}_k^*)\mathbf{x}_k\big)\right] \tag{5.89}$$

$$= E\left[\frac{1}{2}\big(\mathbf{x}_k y_k^* + \mathbf{x}_k^* y_k - \mathbf{x}_k\mathbf{x}_k^H\mathbf{w}_{k-1} - \mathbf{x}_k^*\mathbf{x}_k^T\mathbf{w}_{k-1}\big)\right] \tag{5.90}$$

$$= E\left[\frac{1}{2}(\mathbf{x}_k y_k^* + \mathbf{x}_k^* y_k)\right] - E\left[\frac{1}{2}\big(\mathbf{x}_k\mathbf{x}_k^H + \mathbf{x}_k^*\mathbf{x}_k^T\big)\mathbf{w}_{k-1}\right] \tag{5.91}$$

Since this analysis is done for sufficiently large k, we rename $\mathbf{w}_{k-1}$ to $\mathbf{w}_\infty$ at this point. We further assume that $E[\mathbf{x}_k\mathbf{x}_k^H\mathbf{w}_\infty] = E[\mathbf{x}_k\mathbf{x}_k^H]E[\mathbf{w}_\infty] = E[\mathbf{x}_k\mathbf{x}_k^H]\mathbf{w}_\infty$. Then, (5.91) reads as

$$E[\mathrm{Re}\{e_k\mathbf{x}_k^*\}] = \frac{1}{2}\big(\mathbf{r}_{\mathbf{x}y} + \mathbf{r}_{\mathbf{x}y}^*\big) - \frac{1}{2}(\mathbf{R}_{\mathbf{xx}} + \mathbf{R}_{\mathbf{xx}}^*)\mathbf{w}_\infty \tag{5.92}$$

$$= \mathrm{Re}\{\mathbf{r}_{\mathbf{x}y}\} - \mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}\mathbf{w}_\infty. \tag{5.93}$$

By setting this result equal to zero, the convergence solution immediately follows as

$$\mathbf{w}_\infty = (\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\})^{-1}\mathrm{Re}\{\mathbf{r}_{\mathbf{x}y}\}, \tag{5.94}$$

which corresponds to the optimal Wiener filter in (5.70).

Convergence in the Mean

In this section, we assume $\mathbf{w}_{k-1}$ is independent of the data vector $\mathbf{x}_k$ such as it is done in [72]. Furthermore, we assume $\{y_k, x_k\}$ is independent of $\{y_l, x_l\}$ for $k \neq l$.

We now introduce vector $\mathbf{v}_{k-1} = \mathbf{w}_{k-1} - \mathbf{w}$ as the error vector between the current filter coefficients and the convergence solution. Consider the update equation given in (5.85). Subtracting $\mathbf{w}$ on both sides yields

$$\mathbf{v}_k = \mathbf{v}_{k-1} + \mu\mathrm{Re}\{e_k\mathbf{x}_k^*\} \tag{5.95}$$

$$= \mathbf{v}_{k-1} + \frac{\mu}{2}(\mathbf{x}_k e_k^* + \mathbf{x}_k^* e_k) \tag{5.96}$$

$$= \mathbf{v}_{k-1} + \frac{\mu}{2}\Big(\mathbf{x}_k\big(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\big)^* + \mathbf{x}_k^*\big(y_k - \mathbf{w}_{k-1}^T\mathbf{x}_k\big)\Big) \tag{5.97}$$

$$= \mathbf{v}_{k-1} + \frac{\mu}{2}\big(\mathbf{x}_k y_k^* - \mathbf{x}_k\mathbf{x}_k^H\mathbf{w}_{k-1} + \mathbf{x}_k^* y_k - \mathbf{x}_k^*\mathbf{x}_k^T\mathbf{w}_{k-1}\big). \tag{5.98}$$

The expectation of this expression when utilizing the independence assumptions is given by

$$E[\mathbf{v}_k] = E[\mathbf{v}_{k-1}] + \frac{\mu}{2}\big(\mathbf{r}_{\mathbf{x}y} - \mathbf{R}_{\mathbf{xx}}E[\mathbf{w}_{k-1}] + \mathbf{r}_{\mathbf{x}y}^* - \mathbf{R}_{\mathbf{xx}}^*E[\mathbf{w}_{k-1}]\big) \tag{5.99}$$

$$= E[\mathbf{v}_{k-1}] + \mu(\mathrm{Re}\{\mathbf{r}_{\mathbf{x}y}\} - \mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}E[\mathbf{w}_{k-1}]) \tag{5.100}$$

Utilizing the optimal Wiener filter in (5.70) allows replacing $\mathrm{Re}\{\mathbf{r}_{\mathbf{x}y}\}$ with $\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}\mathbf{w}$, yielding

$$\begin{aligned} E[\mathbf{v}_k] &= E[\mathbf{v}_{k-1}] + \mu(\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}\mathbf{w} - \mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}E[\mathbf{w}_{k-1}]) && (5.101)\\ &= E[\mathbf{v}_{k-1}] - \mu\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}(E[\mathbf{w}_{k-1}] - \mathbf{w}) && (5.102)\\ &= E[\mathbf{v}_{k-1}] - \mu\underbrace{\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}}_{\mathbf{B}}E[\mathbf{v}_{k-1}] && (5.103)\\ &= (\mathbf{I} - \mu\mathbf{B})E[\mathbf{v}_{k-1}], && (5.104) \end{aligned}$$

where $\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\} = \mathbf{B} \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$ is a symmetric matrix. Assuming $\mathbf{B}$ is positive semidefinite, it can be diagonalized by an unitary matrix

$$\mathbf{B} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \tag{5.105}$$

where the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$ contains the real-valued eigenvalues of $\mathbf{B}$ and where $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Inserting (5.105) into (5.104) and introducing the vector $\mathbf{v}' = \mathbf{Q}^T\mathbf{v}$ yields

$$\begin{aligned} E[\mathbf{v}_k] &= \left(\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\right)E[\mathbf{v}_{k-1}] && (5.106)\\ E[\mathbf{v}'_k] &= (\mathbf{I} - \mu\mathbf{\Lambda})E[\mathbf{v}'_{k-1}]. && (5.107) \end{aligned}$$

Since $(\mathbf{I} - \mu\mathbf{\Lambda})$ is a diagonal matrix, the rotated error vectors $\mathbf{v}'_k$ decrease on average if $|1 - \mu\lambda_n| < 1$ holds for $1 \leq n \leq N_\mathbf{w}$, where $\lambda_n \in \mathbb{R}$ denotes the n^{th} eigenvalue of $\mathbf{B}$. This inequality directly leads to

$$0 < \mu < \frac{2}{\lambda_{\max}}, \tag{5.108}$$

where $\lambda_{\max} \in \mathbb{R}$ is the largest eigenvalue of $\mathbf{B}$.

Finally, all results from this section are summarized in

Result 5.1 (LMS Algorithm for Real-Valued Filter Coefficients)

If the adaptive filter is embedded in a complex-valued environment (complex-valued x_k, y_k), but it is known that the optimal filter coefficients shall be real-valued, then the LMS algorithm that produces real-valued estimates $\mathbf{w}_k \in \mathbb{R}^{N_\mathbf{w}}$ is given by:

Initialization:
Initialize $\mathbf{w}_0 = \mathbf{0}^{N_\mathbf{w} \times 1}$;
for $k = 1, 2, \ldots$ **do**
Update $\mathbf{x}_k$ according to (5.2);
Choose step-size μ in accordance with (5.108);
Derive $e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k$;
Evaluate the filter coefficients according to $\mathbf{w}_k = \mathbf{w}_{k-1} + \mu \mathrm{Re}\{e_k \mathbf{x}_k^*\}$;
end

The algorithm converges towards $\mathbf{w} = \left(\mathrm{Re}\{\mathbf{R}_{\mathbf{xx}}\}\right)^{-1} \mathrm{Re}\{\mathbf{r}_{\mathbf{x}y}\}$ in the mean when μ is chosen accordingly, and it is of linear complexity $\mathcal{O}(N_\mathbf{w})$. A detailed complexity analysis can be found in Appendix AE.

5.2.2 RLS Algorithm for Real-Valued Filter Coefficients

A similar extension as it was derived for the LMS algorithm in the previous section is now investigated for the RLS algorithm. We derive an RLS algorithm that incorporates the fact that the coefficients shall be real-valued while the input samples x_k and the desired signal y_k shall be complex-valued.

We start by noticing that the RLS cost function in (5.26) is real-valued even for complex-valued y_i and $\mathbf{x}_i$. Hence, for real-valued $\mathbf{w}_k$ it can be rewritten as

$$J = \sum_{i=0}^{k} \lambda^{k-i} \frac{1}{2} \Big[\big(y_i - \mathbf{w}_k^T \mathbf{x}_i\big)\big(y_i - \mathbf{w}_k^T \mathbf{x}_i\big)^* + \big(y_i - \mathbf{w}_k^T \mathbf{x}_i\big)^* \big(y_i - \mathbf{w}_k^T \mathbf{x}_i\big) \Big] \tag{5.109}$$

$$= \sum_{i=0}^{k} \lambda^{k-i} \frac{1}{2} \Big[y_i y_i^* - \mathbf{w}_k^T \mathbf{x}_i y_i^* - y_i \mathbf{w}_k^T \mathbf{x}_i^* + \mathbf{w}_k^T \mathbf{x}_i \mathbf{x}_i^H \mathbf{w}_k$$

$$+ y_i^* y_i - \mathbf{w}_k^T \mathbf{x}_i^* y_i - y_i^* \mathbf{w}_k^T \mathbf{x}_i + \mathbf{w}_k^T \mathbf{x}_i^* \mathbf{x}_i^T \mathbf{w}_k \Big] \tag{5.110}$$

$$= \sum_{i=0}^{k} \lambda^{k-i} \frac{1}{2} \Big[2 y_i y_i^* - 2\mathbf{w}_k^T \mathbf{x}_i y_i^* - 2\mathbf{w}_k^T \mathbf{x}_i^* y_i + \mathbf{w}_k^T \big(\mathbf{x}_i \mathbf{x}_i^H + \mathbf{x}_i^* \mathbf{x}_i^T\big) \mathbf{w}_k \Big]. \tag{5.111}$$

The partial derivative of (5.111) w.r.t. $\mathbf{w}_k$ follows as

$$\frac{\partial J}{\partial \mathbf{w}_k} = \sum_{i=0}^{k} \lambda^{k-i}\big(-y_i^* \mathbf{x}_i^T - y_i \mathbf{x}_i^H + \mathbf{w}_k^T\big(\mathbf{x}_i \mathbf{x}_i^H + \mathbf{x}_i^* \mathbf{x}_i^T\big)\big) \tag{5.112}$$

$$= \mathbf{w}_k^T \sum_{i=0}^{k} \lambda^{k-i}\big(\mathbf{x}_i \mathbf{x}_i^H + \mathbf{x}_i^* \mathbf{x}_i^T\big) - \sum_{i=0}^{k} \lambda^{k-i}\big(y_i \mathbf{x}_i^H + y_i^* \mathbf{x}_i^T\big). \tag{5.113}$$

Using the definitions in (5.32) and (5.33) allows for

$$\frac{\partial J}{\partial \mathbf{w}_k} = \mathbf{w}_k^T(\mathbf{R}_k + \mathbf{R}_k^*) - \big(\mathbf{r}_k^H + \mathbf{r}_k^T\big). \tag{5.114}$$

Setting (5.114) equal to zero results in

$$\mathbf{w}_k^T = \big(\mathbf{r}_k^H + \mathbf{r}_k^T\big)(\mathbf{R}_k + \mathbf{R}_k^*)^{-1}. \tag{5.115}$$

In order to find a recursive solution, we insert the recursive definitions of $\mathbf{r}_k$ and $\mathbf{R}_k$ in (5.37) and (5.40), respectively, into (5.115), such that

$$\mathbf{w}_k^T = \big(y_k \mathbf{x}_k^H + \lambda \mathbf{r}_{k-1}^H + y_k^* \mathbf{x}_k^T + \lambda \mathbf{r}_{k-1}^T\big)\big(\mathbf{x}_k \mathbf{x}_k^H + \lambda \mathbf{R}_{k-1} + \mathbf{x}_k^* \mathbf{x}_k^T + \lambda \mathbf{R}_{k-1}^*\big)^{-1}. \tag{5.116}$$

By utilizing the notations

$$\underline{\mathbf{y}}_k = \begin{bmatrix} y_k \\ y_k^* \end{bmatrix} \tag{5.117}$$

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{x}_k & \mathbf{x}_k^* \end{bmatrix} \tag{5.118}$$

$$\widetilde{\mathbf{r}}_k = \mathbf{r}_k + \mathbf{r}_k^* \tag{5.119}$$

$$\widetilde{\mathbf{R}}_k = \mathbf{R}_k + \mathbf{R}_k^*, \tag{5.120}$$

(5.116) can be reformulated as

$$\mathbf{w}_k^T = \Big(\underline{\mathbf{y}}_k^T \mathbf{X}_k^H + \lambda \widetilde{\mathbf{r}}_{k-1}^H\Big)\Big(\mathbf{X}_k \mathbf{X}_k^H + \lambda \widetilde{\mathbf{R}}_{k-1}\Big)^{-1}. \tag{5.121}$$

Note that by comparing (5.115) with (5.121) it immediately follows that

$$\widetilde{\mathbf{r}}_k^H = \mathbf{r}_k^H + \mathbf{r}_k^T = \underline{\mathbf{y}}_k^T \mathbf{X}_k^H + \lambda \widetilde{\mathbf{r}}_{k-1}^H \tag{5.122}$$

and

$$\widetilde{\mathbf{R}}_k = \mathbf{X}_k \mathbf{X}_k^H + \lambda \widetilde{\mathbf{R}}_{k-1}. \tag{5.123}$$

Applying Woodbury's matrix inversion lemma [54] to derive the inverse of $\widetilde{\mathbf{R}}_k$ produces

$$\widetilde{\mathbf{R}}_k^{-1} = \Big(\lambda \widetilde{\mathbf{R}}_{k-1}\Big)^{-1} - \Big(\lambda \widetilde{\mathbf{R}}_{k-1}\Big)^{-1} \mathbf{X}_k \Big(\mathbf{I}^{2\times 2} + \mathbf{X}_k^H \Big(\lambda \widetilde{\mathbf{R}}_{k-1}\Big)^{-1} \mathbf{X}_k\Big)^{-1} \mathbf{X}_k^H \Big(\lambda \widetilde{\mathbf{R}}_{k-1}\Big)^{-1}. \tag{5.124}$$

By introducing the notation

$$\widetilde{\mathbf{P}}_k = \widetilde{\mathbf{R}}_k^{-1} \tag{5.125}$$

Eq. (5.124) follows as

$$\widetilde{\mathbf{P}}_k = \lambda^{-1}\widetilde{\mathbf{P}}_{k-1} - \lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big)^{-1}\mathbf{X}_k^H\lambda^{-1}\widetilde{\mathbf{P}}_{k-1}. \tag{5.126}$$

We now introduce the *gain matrix* as

$$\mathbf{G}_k = \lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big)^{-1} \tag{5.127}$$

such that (5.126) becomes more compact

$$\begin{aligned}\widetilde{\mathbf{P}}_k &= \lambda^{-1}\widetilde{\mathbf{P}}_{k-1} - \mathbf{G}_k\mathbf{X}_k^H\lambda^{-1}\widetilde{\mathbf{P}}_{k-1} && (5.128)\\ &= \lambda^{-1}\Big(\widetilde{\mathbf{P}}_{k-1} - \mathbf{G}_k\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\Big). && (5.129)\end{aligned}$$

Since $\widetilde{\mathbf{P}}_k = \widetilde{\mathbf{R}}_k^{-1}$ is a Hermitian matrix according to (5.120), (5.129) can be expressed as

$$\widetilde{\mathbf{P}}_k = \lambda^{-1}\Big(\widetilde{\mathbf{P}}_{k-1} - \widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\mathbf{G}_k^H\Big). \tag{5.130}$$

Furthermore, a reformulation of (5.127) yields

$$\begin{aligned}\mathbf{G}_k\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big) &= \lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k && (5.131)\\ \mathbf{G}_k + \mathbf{G}_k\lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k &= \lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k && (5.132)\\ \mathbf{G}_k &= \lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k - \mathbf{G}_k\lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k && (5.133)\\ \mathbf{G}_k &= \lambda^{-1}\Big(\widetilde{\mathbf{P}}_{k-1} - \mathbf{G}_k\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\Big)\mathbf{X}_k. && (5.134)\end{aligned}$$

The right hand side of (5.134) can be identified as $\widetilde{\mathbf{P}}_k\mathbf{X}_k$ according to (5.129) such that

$$\mathbf{G}_k = \widetilde{\mathbf{P}}_k\mathbf{X}_k. \tag{5.135}$$

We are now able to find an expression for $\mathbf{w}_k$ in (5.115)

$$\begin{aligned}\mathbf{w}_k^T &= \big(\mathbf{r}_k^H + \mathbf{r}_k^T\big)(\mathbf{R}_k + \mathbf{R}_k^*)^{-1} && (5.136)\\ &= \widetilde{\mathbf{r}}_k^H\widetilde{\mathbf{P}}_k. && (5.137)\end{aligned}$$

Incorporating (5.122) and (5.135) into (5.137) produces

$$\begin{aligned}\mathbf{w}_k^T &= \underline{\mathbf{y}}_k^T\mathbf{X}_k^H\widetilde{\mathbf{P}}_k + \lambda\widetilde{\mathbf{r}}_{k-1}^H\widetilde{\mathbf{P}}_k && (5.138)\\ &= \lambda\widetilde{\mathbf{r}}_{k-1}^H\widetilde{\mathbf{P}}_k + \underline{\mathbf{y}}_k^T\mathbf{G}_k^H. && (5.139)\end{aligned}$$

Combining the recursive definition of $\widetilde{\mathbf{P}}_k$ in (5.130) with (5.139) allows to express

$$\begin{aligned}\mathbf{w}_k^T &= \widetilde{\mathbf{r}}_{k-1}^H\Big(\widetilde{\mathbf{P}}_{k-1} - \widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\mathbf{G}_k^H\Big) + \underline{\mathbf{y}}_k^T\mathbf{G}_k^H && (5.140)\\ &= \mathbf{w}_{k-1}^T - \mathbf{w}_{k-1}^T\mathbf{X}_k\mathbf{G}_k^H + \underline{\mathbf{y}}_k^T\mathbf{G}_k^H && (5.141)\\ &= \mathbf{w}_{k-1}^T + \Big(\underline{\mathbf{y}}_k^T - \mathbf{w}_{k-1}^T\mathbf{X}_k\Big)\mathbf{G}_k^H. && (5.142)\end{aligned}$$

The term $\mathbf{w}_{k-1}^T \mathbf{X}_k$ can be identified as the augmented output $\hat{\underline{\mathbf{y}}}_k^T$ of the filter such that

$$\underline{\mathbf{y}}_k^T - \mathbf{w}_{k-1}^T \mathbf{X}_k = \underline{\mathbf{y}}_k^T - \hat{\underline{\mathbf{y}}}_k^T \tag{5.143}$$
$$= \underline{\mathbf{e}}_k^T, \tag{5.144}$$

where $\underline{\mathbf{e}}_k$ represents the augmented a-priori error. Reinserting (5.144) into (5.142) yields

$$\mathbf{w}_k^T = \mathbf{w}_{k-1}^T + \underline{\mathbf{e}}_k^T \mathbf{G}_k^H. \tag{5.145}$$

This result represents the update equation from the old filter coefficients in $\mathbf{w}_{k-1}$ to the new ones in $\mathbf{w}_k$. We propose to use the same initialization as for the ordinary RLS algorithm, which leads to

Result 5.2 (RLS Algorithm for Real-Valued Filter Coefficients)

If the adaptive filter is embedded in a complex-valued environment (complex-valued x_k, y_k), but it is known that the optimal filter coefficients shall be real-valued, then the RLS algorithm that produces real-valued estimates $\mathbf{w}_k \in \mathbb{R}^{N_\mathbf{w}}$ is given by:

Initialization:
Choose λ and δ;
$\widetilde{\mathbf{P}}_0 = \delta \mathbf{I}^{N_\mathbf{w} \times N_\mathbf{w}}$;
$\mathbf{w}_0 = \mathbf{0}^{N_\mathbf{w} \times 1}$;
for $k = 1, 2, \ldots$ **do**
Update $\mathbf{x}_k \in \mathbb{C}^{N_\mathbf{w}}$ according to (5.2);
Construct $\mathbf{X}_k \in \mathbb{C}^{N_\mathbf{w} \times 2}$ according to (5.118);
Construct $\underline{\mathbf{y}}_k \in \mathbb{C}^2$ according to (5.117);
Derive $\mathbf{G}_k \in \mathbb{C}^{N_\mathbf{w} \times 2}$ according to (5.127);
Derive $\underline{\mathbf{e}}_k \in \mathbb{R}^2$ according to (5.144);
Update $\widetilde{\mathbf{P}}_k \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$ according to (5.129);
Evaluate the new filter coefficients $\mathbf{w}_k \in \mathbb{R}^{N_\mathbf{w}}$ according to (5.145);
end

The algorithm is of quadratic complexity $\mathcal{O}(N_\mathbf{w}^2)$. A detailed complexity analysis can be found in Appendix AF.

Note that the final algorithm contains a matrix inversion in (5.127). However, this matrix is only of size 2×2, which is trivial to invert.

For the LMS algorithm for real-valued filter coefficients, it turned out that the optimal update equation of $\mathbf{w}_k$ corresponds to the intuitive approach of taking only the real values of $e_k \mathbf{x}_k^*$ to update $\mathbf{w}_k$. In contrast to that, the derived RLS algorithm for real-valued filter coefficients utilizes a different update equation than the intuitive approach based on the ordinary RLS algorithm discussed in Exampe 5.1. The performance gain achievable with these derived algorithms is presented in the next simulation example.

Example 5.1 (Estimation of a Real-Valued Impulse Response With Adaptive Filters)

The example considers the task of system identification according to Figure 5.1 with the additional knowledge that the impulse response of the unknown system is real-valued. The true real-valued impulse responses $\mathbf{h} \in \mathbb{R}^{N_{\mathbf{h}}}$ with length $N_{\mathbf{h}} = 5$ were randomly drawn from a zero mean Gaussian distribution with covariance matrix $\mathbf{C}_{\mathbf{hh}} = \mathbf{I}^{5\times5}$. The complex-valued input and noise samples were drawn from zero mean complex proper Gaussian distributions with variances 1 and 10^{-4}, respectively. All considered adaptive filters utilize $N_{\mathbf{w}} = N_{\mathbf{h}}$ and are listed in the following:

1. The ordinary LMS algorithm derived in Section 5.1.
2. The LMS algorithm for real-valued filter coefficients from Result 5.1.
3. The ordinary RLS algorithm derived in Section 5.1.
4. The intuitive algorithm resulting from the ordinary RLS algorithm when replacing the update equation in (5.59) by

$$\mathbf{w}_k^T = \mathbf{w}_{k-1}^T + \mathrm{Re}\{e_k \mathbf{g}_k^H\} \tag{5.146}$$

   resulting in real-valued vectors $\mathbf{w}_k$.

5. The RLS algorithm for real-valued filter coefficients from Result 5.2.

The LMS based algorithms utilize $\mu = 0.05$ and the RLS based algorithms utilize $\lambda = 0.99$ and $\delta = 100$. The convergence curves in terms of $E_{\mathbf{h},x,n}[||\mathbf{w}_k - \mathbf{h}||^2]$ are presented in Figure 5.2.

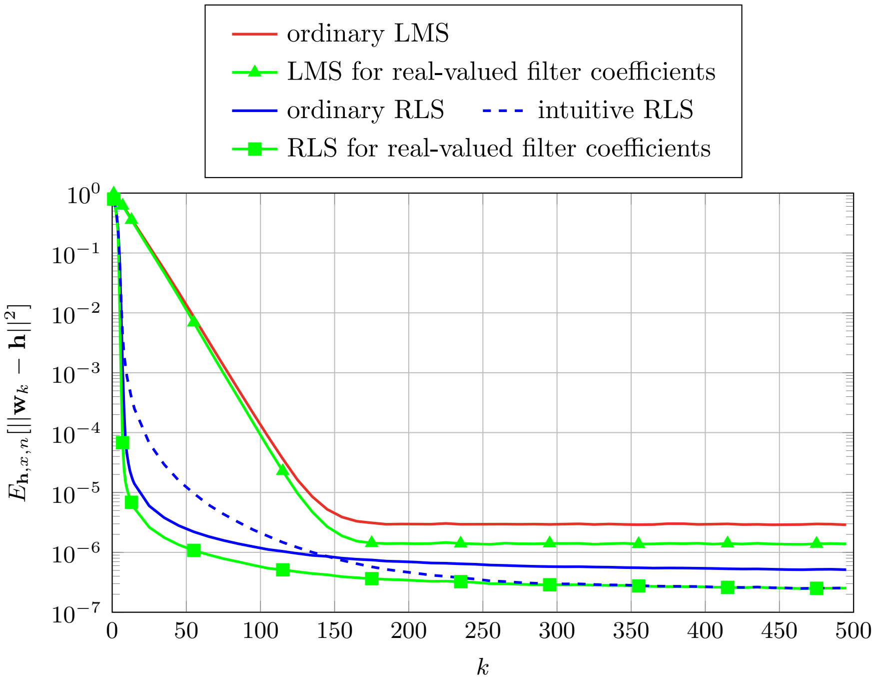

Figure 5.2: Convergence curves of various adaptive filters. The complex-valued input and noise samples were drawn from zero mean complex proper Gaussian distributions with variances 1 and 10^{-4}, respectively. The true real-valued impulse responses were randomly drawn from a zero mean Gaussian distribution with covariance matrix $\mathbf{C}_{\mathbf{hh}} = \mathbf{I}^{5\times5}$.

The discussion starts with the LMS based filters. The LMS algorithm for real-valued filter coefficients yields a slightly faster convergence speed and an increased steady state performance compared to the ordinary LMS algorithm. Similar to that, the RLS algorithm for real-valued filter coefficients also outperforms the ordinary RLS algorithm. Moreover, it significantly beats the intuitive RLS algorithm in terms of the convergence speed.

A second example from a real-world application in the context of wireless transceivers is presented in the following.

Example 5.2 (Transmitter Leakage Cancellation)

Modern wireless transceivers employ frequency division duplex operation where both, the transmit path (TX) as well as the receive path (RX), are active at the same time. Due to a limited TX-RX isolation of the duplexer, the TX signal leaks into the RX path. This leakage signal, although being at a different frequency than the RX signal, can lead to a baseband interference in the receiver due to the downconversion by an unwanted spur [16,74,75]. Therefore, the extraction and cancellation of the leakage signal has to be targeted [16,74,75]. One way to do this is to implement an additional so called *auxiliary receive path*, which shall only receive the leakage signal [16]. A simplified baseband equivalent model is shown in Figure 5.3.

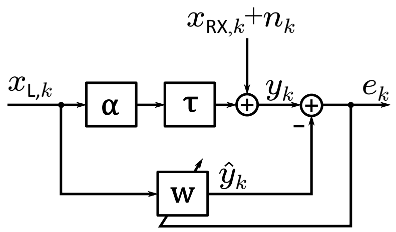

Figure 5.3: System model for the adaptive filter application in a modern wireless transceiver. $\alpha \in \mathbb{R}$ represents a gain and $\tau \in \mathbb{R}$ denotes a fractional delay.

The upper path reflects the main receiver containing the leakage signal $x_{\mathrm{L},k}$, the wanted receive signal $x_{\mathrm{RX},k}$ and noise n_k. The lower path, reflecting the auxiliary receiver, only contains the leakage signal. However, due to non-idealities of the analog circuits, the two paths typically have different gains and delays. This is incorporated in the model by introducing a gain α and a fractional delay τ in the main path. The task of the adaptive filter is to estimate the delayed and amplified version of $x_{\mathrm{L},k}$. Finally, e_k shall match $n_k + x_{\mathrm{RX},k}$ since the leakage signal is cancelled.

In the following simulation, the gain α is chosen to be 1 for simplicity. The fractional delay shall be 14.3 samples. Note that the fractional delay has been implemented in simulation with the help of an intermediate oversampling stage. It has been shown in [16,76], that such a fractional delay results in approximately sinc-shaped vectors $\mathbf{w}_k$. For the sake of simplicity, the term $x_{\mathrm{RX},k} + n_k$ is approximated as zero mean complex proper Gaussian with variance 0.1. Furthermore, $x_{\mathrm{L},k}$ is approximated as a zero mean complex proper Gaussian with variance 1. Since the real and complex-valued parts of $x_{\mathrm{L},k}$ are attenuated and delayed equally, the optimal filter coefficients in $\mathbf{w}_k$ are real-valued. This allows to apply the derived adaptive filters in Result 5.1 and Result 5.2. Moreover, also the ordinary LMS algorithm, the ordinary RLS algorithm as well as the intuitive adaptive filters introduced in Example 5.1 are implemented for the described cancellation task. The LMS based algorithms utilize $\mu = 0.005$ and the RLS based algorithms utilize $\lambda = 0.999$ and $\delta = 100$.

The overall goal is that the samples e_k in Figure 5.3 match $x_{\mathrm{RX},k} + n_k$. Hence, $E[|e_k - x_{\mathrm{RX},k} - n_k|^2]$ serves as performance measure. These mean square values (averaged over many simulation runs) are shown in Figure 5.4. The performance curves are very similar to that in Example 5.1. Again, the LMS algorithm for real-valued filter coefficients yields a slightly faster convergence speed and an increased steady state performance compared to the ordinary LMS algorithm. Also the RLS algorithm for real-valued filter coefficients clearly outperforms the ordinary RLS as well as the intuitive RLS algorithm.

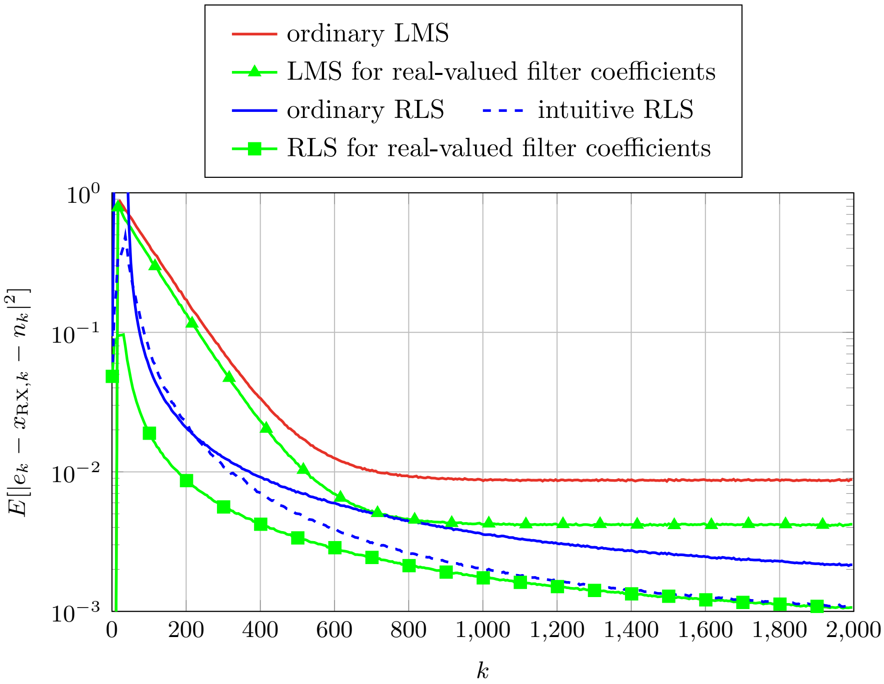

Figure 5.4: Convergence curves of various adaptive filters.

5.3 Adaptive Filters incorporating Prior Knowledge

In many system identification applications, statistics about the impulse response of the unknown system that shall be estimated are available. These statistics are usually termed *prior knowledge*, especially in the context of Bayesian estimation (cf. Chapter 4). In this context, the impulse response $\mathbf{h}$ is a random variable whose particular realization has to be estimated.

Now, we assume statistics about the impulse response $\mathbf{h}$ are available in form of the mean vector $E_{\mathbf{h}}[\mathbf{h}]$ and the positive definite covariance matrix $\mathbf{C}_{\mathbf{hh}}$. The goal is to estimate the particular realization of $\mathbf{h}$ with the help of adaptive filters that incorporate the knowledge about $E_{\mathbf{h}}[\mathbf{h}]$ and $\mathbf{C}_{\mathbf{hh}}$.

An algorithm that is related to the RLS algorithm and that uses $E_{\mathbf{h}}[\mathbf{h}]$ and $\mathbf{C}_{\mathbf{hh}}$ is the sequential LMMSE estimator. Its implementation in the context of the discussed adaptive filter scenario is given by

Initialization:
  $\mathbf{P}_0 = \mathbf{C}_{\mathbf{hh}}$;
  $\mathbf{w}_0 = E_{\mathbf{h}}[\mathbf{h}]$;
for $k = 1, 2, \ldots$ **do**
  Update $\mathbf{x}_k$ according to (5.2);
  Derive e_k according to: $e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k$;
  Determine $\mathbf{g}_k$ according to: $\mathbf{g}_k = \mathbf{P}_{k-1}\mathbf{x}_k^* \left(\sigma_n^2 + \mathbf{x}_k^T \mathbf{P}_{k-1} \mathbf{x}_k^*\right)^{-1}$;
  Derive $\mathbf{P}_k$ according to: $\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{g}_k \mathbf{x}_k^T \mathbf{P}_{k-1}$;
  Evaluate the new filter coefficients $\mathbf{w}_k$ according to (5.59);
end

The main differences to the RLS algorithm are that prior knowledge is utilized in the filter initialization, that the forgetting factor λ is dismissed, and that the noise variance σ_n^2 is incorporated. Note that the noise variance can be time-variant.

This adaptive filter formulation of the sequential LMMSE estimator incorporates prior knowledge about $\mathbf{h}$. Hence, an RLS-type adaptive filter incorporating prior knowledge about $\mathbf{h}$ already exists. However, to the best of our knowledge, a similar extension for the LMS algorithm has not been existing in literature. We provided such an extension in [77], whose derivation is repeated in the following. These investigations can be considered to be relevant for many applications since the LMS algorithm is more widespread than the RLS algorithm.

Bayesian LMS Algorithm

In the derivation of the so called Bayesian LMS algorithm we assume that m measurements are conducted (typically during a training phase). Let the measurements be put together in the vector

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{C}^m. \tag{5.147}$$

The derivation is based on the *posterior* PDF denoted as $p(\mathbf{h}|\mathbf{y})$. This PDF describes the probability density of the unknown impulse response $\mathbf{h}$ given all the measurements. For now m shall be finite, later we consider a potentially infinite number of measurements. According to Bayes' rule [78], the posterior PDF can be rewritten as

$$\begin{aligned} p(\mathbf{h}|\mathbf{y}) &= p(\mathbf{y}|\mathbf{h})p(\mathbf{h})/p(\mathbf{y}) &\quad (5.148)\\ &\propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}), &\quad (5.149)\end{aligned}$$

where $p(\mathbf{y}|\mathbf{h})$ and $p(\mathbf{h})$ are denoted as the *likelihood* PDF and the *prior* PDF, respectively. Note that the MAP estimator for $\mathbf{h}$ is the vector maximizing the posterior PDF, i.e.

$$\hat{\mathbf{h}}_{\text{MAP}} = \arg \max_{\mathbf{h}} \; p(\mathbf{y}|\mathbf{h})p(\mathbf{h}) \tag{5.150}$$

$$= \arg \max_{\mathbf{h}} \; \log \; (p(\mathbf{y}|\mathbf{h})p(\mathbf{h})), \tag{5.151}$$

where 'log' denotes the natural logarithm. However, we are more interested in an iterative solution for the MAP estimator. For this, we make further modifications of (5.151). The connection between the vector of measurements $\mathbf{y}$ and the impulse response $\mathbf{h}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{h} + \mathbf{n}, \tag{5.152}$$

where the rows of $\mathbf{H} \in \mathbb{C}^{m \times N_\mathbf{w}}$ are provided by the input samples $\mathbf{x}_i$ in (5.2), such that

$$\mathbf{H} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_m^T \end{bmatrix}. \tag{5.153}$$

Assuming the noise $\mathbf{n}$ in (5.152) to be complex proper Gaussian distributed with zero mean and with known covariance matrix $\mathbf{C}_{\mathbf{nn}}$, the likelihood PDF directly follows as

$$p(\mathbf{y}|\mathbf{h}) = \frac{1}{\pi^m \det(\mathbf{C}_{\mathbf{nn}})} \mathrm{e}^{-(\mathbf{y}-\mathbf{H}\mathbf{h})^H \mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{h})}. \tag{5.154}$$

The next assumption is, that the prior PDF is also complex proper Gaussian according to

$$p(\mathbf{h}) = \frac{1}{\pi^{N_\mathbf{w}} \det(\mathbf{C}_{\mathbf{hh}})} \mathrm{e}^{-(\mathbf{h}-E_\mathbf{h}[\mathbf{h}])^H \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h}-E_\mathbf{h}[\mathbf{h}])}, \tag{5.155}$$

where $E_\mathbf{h}[\mathbf{h}]$ and $\mathbf{C}_{\mathbf{hh}}$ are the known mean vector and covariance matrix of $\mathbf{h}$. Inserting (5.154) and (5.155) into (5.151) produces

$$\hat{\mathbf{h}}_{\text{MAP}} = \arg \max_{\mathbf{h}} \; \Big(-(\mathbf{y}-\mathbf{H}\mathbf{h})^H \mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{h}) - (\mathbf{h}-E_\mathbf{h}[\mathbf{h}])^H \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h}-E_\mathbf{h}[\mathbf{h}]) \Big) \tag{5.156}$$

$$= \arg \min_{\mathbf{h}} \; \underbrace{(\mathbf{y}-\mathbf{H}\mathbf{h})^H \mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{h}) + (\mathbf{h}-E_\mathbf{h}[\mathbf{h}])^H \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h}-E_\mathbf{h}[\mathbf{h}])}_{J(\mathbf{h})} \tag{5.157}$$

$$= \arg \min_{\mathbf{h}} \; J(\mathbf{h}). \tag{5.158}$$

To obtain an iterative solution of the MAP estimator in (5.158) a gradient-based approach is used. The derivative of $J(\mathbf{h})$ w.r.t. $\mathbf{h}^*$ is

$$\frac{\partial J(\mathbf{h})}{\partial \mathbf{h}^*} = -\mathbf{y}^T \big(\mathbf{C}_{\mathbf{nn}}^{-1}\big)^T \mathbf{H}^* + \mathbf{h}^T \mathbf{H}^T \big(\mathbf{C}_{\mathbf{nn}}^{-1}\big)^T \mathbf{H}^* + \mathbf{h}^T \big(\mathbf{C}_{\mathbf{hh}}^{-1}\big)^T - E_\mathbf{h}[\mathbf{h}]^T \big(\mathbf{C}_{\mathbf{hh}}^{-1}\big)^T. \tag{5.159}$$

Replacing $\mathbf{h}$ with the estimate $\mathbf{w}_{k-1}$ and inserting (5.159) into the update equation in (5.9) produces

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu\big(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{H}\mathbf{w}_{k-1} - \mathbf{y}) + \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}])\big). \tag{5.160}$$

This update equation (with appropriately chosen μ) converges to the MAP estimator for the case of complex proper Gaussian likelihood and prior PDFs. In the following reformulations and simplifications, we assume the noise covariance matrix to be a scaled identity matrix[6] $\mathbf{C}_{\mathbf{nn}} = \sigma_n^2 \mathbf{I}$. Recall that the i^{th} row of $\mathbf{H}$ is given by $\mathbf{x}_i^T$ according to (5.153). With that, the update equation can be reformulated as

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu \sum_{i=1}^{m} \left(\frac{1}{\sigma_n^2} \mathbf{x}_i^* \big(\mathbf{x}_i^T \mathbf{w}_{k-1} - y_i\big) + a_i \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}]) \right). \tag{5.161}$$

The scalars $a_i \in \mathbb{R}$ in (5.161) for $i = 1, \ldots, m$ are arbitrary except for the constraint

$$\sum_{i=1}^{m} a_i = 1. \tag{5.162}$$

Eq. 5.161 allows to use a simplification similar as done for the approximate least squares (ALS) [79] or the Kaczmarz algorithm [80] by using only one of the m partial gradients per iteration, which yields

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \mu \left(\frac{1}{\sigma_n^2} \mathbf{x}_k^* \big(\mathbf{x}_k^T \mathbf{w}_{k-1} - y_k\big) + a_k \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}]) \right). \tag{5.163}$$

Identifying the term $\mathbf{x}_k^T \mathbf{w}_{k-1} - y_k$ in (5.163) as the negative a-priori error e_k leads to an LMS-like algorithm of the form [77]

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu \left(\frac{1}{\sigma_n^2} \mathbf{x}_k^* e_k - a_k \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}]) \right). \tag{5.164}$$

The Bayesian LMS algorithm in (5.164) is an LMS-based adaptive filter that incorporates prior knowledge about the unknown impulse response in form of $E_{\mathbf{h}}[\mathbf{h}]$ and $\mathbf{C}_{\mathbf{hh}}$. The term $\mathbf{x}_k^* e_k$ can easily be identified as the update term of the ordinary LMS algorithm in (5.19). The Bayesian LMS algorithm scales this term by $\frac{1}{\sigma_n^2}$ and adds a term that incorporates the prior knowledge. Now, the scalars a_k must fulfill

$$\sum_{i=1}^{m} a_k = 1, \tag{5.165}$$

where m is the total number of measurements ($1 \leq k \leq m$). A possible choice for a_k is

$$a_k = \frac{1}{m} \tag{5.166}$$

which requires the knowledge of the total number of measurements m in advance. For the theoretical case of an infinite number of measurements

$$a_k = \frac{1}{2^k}, \tag{5.167}$$

[6] A diagonal noise covariance matrix is also possible with a slight modification of the derivation.

is a possible option. Note that (5.166) can also be employed for the theoretical case of an infinite number of measurements. Then, m marks a time index where $a_k = 0$ for $k > m$. In this case, (5.164) reduces to the ordinary LMS update equation for $k > m$ except for a scalar term that can be moved into the step-size. In order to decrease the computational complexity of the algorithm, only (5.166) is considered in the following. As initialization, $\mathbf{x}_0 = E_{\mathbf{h}}[\mathbf{h}]$ is employed in accordance with the sequential LMMSE.

Prior knowledge helps to speed up the convergence in the mean, but for large k, the algorithm reaches the same performance as the usual LMS algorithm since then, the measurements dominate the prior information about $\mathbf{h}$. Hence, the algorithm in (5.164) converges to the same filter coefficients in the mean as the ordinary LMS algorithm for $k \to \infty$ for two reasons:

- the measurements dominate the prior information about $\mathbf{h}$ for large k,
- a_k is assumed to be zero for large enough k also in the case of an infinite number of measurements.

The convergence properties are demonstrated in Example 5.3 later on.

Convergence in the Mean

We recently stated that the algorithm in (5.164) converges to the same filter coefficients in the mean as the ordinary LMS algorithm for $k \to \infty$. However, in order to derive a normalized version of the Bayesian LMS algorithm, we perform a convergence analysis. To make the following investigations mathematically tractable, we assume $\mathbf{w}_k$ is independent of the data vector $\mathbf{x}_k$ such as it is done in [72]. Furthermore, we assume $\{y_k, x_k\}$ is independent of $\{y_l, x_l\}$ for $k \neq l$. The investigations will lead to lower and upper bounds for the step size μ.

The error between y_k and $\hat{y}_k$ is given by

$$
\begin{aligned}
e_k &= y_k - \hat{y}_k && (5.168)\\
&= \mathbf{h}^T \mathbf{x}_k + n_k - \mathbf{w}_{k-1}^T \mathbf{x}_k && (5.169)\\
&= \underbrace{\left(\mathbf{h}^T - \mathbf{w}_{k-1}^T\right)}_{\mathbf{v}_{k-1}^T} \mathbf{x}_k + n_k && (5.170)\\
&= \mathbf{v}_{k-1}^T \mathbf{x}_k + n_k, && (5.171)
\end{aligned}
$$

where $\mathbf{h} - \mathbf{w}_{k-1} = \mathbf{v}_{k-1} \in \mathbb{R}^{N_{\mathbf{w}}}$. Inserting (5.171) into the update equation in (5.164) yields

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu\bigg(\frac{1}{\sigma_n^2}\mathbf{x}_k^*\mathbf{v}_{k-1}^T\mathbf{x}_k + \frac{1}{\sigma_n^2}\mathbf{x}_k^* n_k - a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}])\bigg) \tag{5.172}$$

$$= \mathbf{w}_{k-1} + \mu\bigg(\frac{1}{\sigma_n^2}\mathbf{x}_k^*\mathbf{x}_k^T\mathbf{v}_{k-1} + \frac{1}{\sigma_n^2}\mathbf{x}_k^* n_k - a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h} - \mathbf{v}_{k-1} - E_{\mathbf{h}}[\mathbf{h}])\bigg) \tag{5.173}$$

$$= \mathbf{w}_{k-1} + \mu\bigg(\frac{1}{\sigma_n^2}\mathbf{x}_k^*\mathbf{x}_k^T\mathbf{v}_{k-1} + \frac{1}{\sigma_n^2}\mathbf{x}_k^* n_k + a_k\mathbf{C}_{\mathbf{hh}}^{-1}\mathbf{v}_{k-1} - a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h} - E_{\mathbf{h}}[\mathbf{h}])\bigg) \tag{5.174}$$

$$= \mathbf{w}_{k-1} + \mu\bigg(\Big(\frac{1}{\sigma_n^2}\mathbf{x}_k^*\mathbf{x}_k^T + a_k\mathbf{C}_{\mathbf{hh}}^{-1}\Big)\mathbf{v}_{k-1} + \frac{1}{\sigma_n^2}\mathbf{x}_k^* n_k - a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h} - E_{\mathbf{h}}[\mathbf{h}])\bigg). \tag{5.175}$$

Subtracting $\mathbf{h}$ from both sides results in

$$\mathbf{v}_k = \mathbf{v}_{k-1} - \mu\bigg(\Big(\frac{1}{\sigma_n^2}\mathbf{x}_k^*\mathbf{x}_k^T + a_k\mathbf{C}_{\mathbf{hh}}^{-1}\Big)\mathbf{v}_{k-1} + \frac{1}{\sigma_n^2}\mathbf{x}_k^* n_k - a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h} - E_{\mathbf{h}}[\mathbf{h}])\bigg). \tag{5.176}$$

Applying the expectation operator conditioned on $\mathbf{h}$ to both sides of (5.176) and applying the independence assumptions yields

$$E[\mathbf{v}_k|\mathbf{h}] = E[\mathbf{v}_{k-1}|\mathbf{h}] - \mu\underbrace{\Big(\frac{1}{\sigma_n^2}\mathbf{R}_{\mathbf{xx}}^* + a_k\mathbf{C}_{\mathbf{hh}}^{-1}\Big)}_{\mathbf{B}_k}E[\mathbf{v}_{k-1}|\mathbf{h}] + \mu a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h} - E_{\mathbf{h}}[\mathbf{h}]) \tag{5.177}$$

$$= (\mathbf{I} - \mu\mathbf{B}_k)E[\mathbf{v}_{k-1}|\mathbf{h}] + \mu a_k\mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{h} - E_{\mathbf{h}}[\mathbf{h}]), \tag{5.178}$$

where $\frac{1}{\sigma_n^2}\mathbf{R}_{\mathbf{xx}}^* + a_k\mathbf{C}_{\mathbf{hh}}^{-1} = \mathbf{B}_k \in \mathbb{C}^{N_{\mathbf{w}}\times N_{\mathbf{w}}}$. Note that $\mathbf{B}_k$ is hermitian and positive definite since $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{hh}}$ are both hermitian and positive definite. This allows to utilize the same approach for deriving bounds for the step-size as in (5.104)–(5.108). By doing so, conditions for the convergence in the mean of the algorithm are given by

$$0 < \mu_k < \frac{2}{\lambda_{\max}(\mathbf{B}_k)} = \frac{2}{\lambda_{\max}\Big(\frac{1}{\sigma_n^2}\mathbf{R}_{\mathbf{xx}}^* + a_k\mathbf{C}_{\mathbf{hh}}^{-1}\Big)}, \tag{5.179}$$

where $\lambda_{\max}(\mathbf{B}_k)$ denotes the largest eigenvalue of the matrix $\mathbf{B}_k$. Note that the resulting step-size depends on k since $\mathbf{B}_k$ depends on k. We now simplify this expression. For two symmetric $p \times p$ matrices $\mathbf{E}$ and $\mathbf{F}$ it holds that the maximum eigenvalue of the sum of matrices $\lambda_{\max}(\mathbf{E}+\mathbf{F})$ is smaller or equal than the sum of the maximum eigenvalues of the matrices [81]:

$$\lambda_{\max}(\mathbf{E}+\mathbf{F}) \leq \lambda_{\max}(\mathbf{E}) + \lambda_{\max}(\mathbf{F}). \tag{5.180}$$

With that we get

$$0 < \mu_k < \frac{2}{\frac{1}{\sigma_n^2}\lambda_{\max}(\mathbf{R}_{\mathbf{xx}}^*) + a_k\lambda_{\max}\big(\mathbf{C}_{\mathbf{hh}}^{-1}\big)}. \tag{5.181}$$

Approximating $\mathbf{R}_{\mathbf{xx}}^*$ with its instantaneous estimate $\mathbf{x}_k^*\mathbf{x}_k^T$, and incorporating $\lambda_{\max}\big(\mathbf{x}_k^*\mathbf{x}_k^T\big) = \|\mathbf{x}_k\|_2^2$, allows to approximate the upper bound as $\frac{2}{\frac{1}{\sigma_n^2}\|\mathbf{x}_k\|_2^2 + a_k\lambda_{\max}\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)}$.

The update equation in (5.164) in combination with the boundaries for the step-size in derived upper bound motivates the Bayesian NLMS of the form

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu_{n,k} \frac{1}{\epsilon + \frac{1}{\sigma_n^2}\|\mathbf{x}_k\|_2^2 + a_k \lambda_{\max}\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)} \left(\frac{1}{\sigma_n^2} \mathbf{x}_k^* e_k - a_k \mathbf{C}_{\mathbf{hh}}^{-1} (\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}]) \right), \tag{5.182}$$

where the normalized step-size $\mu_{n,k}$ is usually chosen between 0 and 1 for all k. $\epsilon \in \mathbb{R}$ in (5.182) is a small positive-valued constant to overcome possible instabilities when $a_k = 0$ and when $\|\mathbf{x}_k\|_2^2$ is very small.

This Bayesian NLMS is summarized in

Result 5.3 (Bayesian NLMS Algorithm)

Consider the adaptive filtering task described in Section 5.1, where the impulse response $\mathbf{h} \in \mathbb{R}^{N_{\mathbf{w}}}$ is a random variable whose particular realization has to be estimated. If prior knowledge about $\mathbf{h}$ is available in form of its mean vector $E_{\mathbf{h}}[\mathbf{h}]$ and covariance matrix $\mathbf{C}_{\mathbf{hh}}$, then the Bayesian NLMS algorithm that incorporates this prior knowledge is given by:

Initialization:
  Initialize $\mathbf{w}_0 = E_{\mathbf{h}}[\mathbf{h}]$;
  Pre-evaluate $\mathbf{C}_{\mathbf{hh}}^{-1}$;
  Pre-evaluate $\lambda_{\max}\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)$;
  Pre-evaluate $\frac{1}{\sigma_n^2}$;
  Choose $\epsilon \in \mathbb{R}$ and $a_k \in \mathbb{R}$ (e.g., (5.166));
for $k = 0, 1, \ldots$ **do**
  Update $\mathbf{x}_k$ according to (5.2);
  Choose step-size $\mu_{n,k}$ between 0 and 1;
  Derive $e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k$;
  Evaluate the filter coefficients according to (5.182);
end

The algorithm is of quadratic complexity $\mathcal{O}(N_{\mathbf{w}}^2)$ for general covariance matrices $\mathbf{C}_{\mathbf{hh}}$. However, for the case of diagonal $\mathbf{C}_{\mathbf{hh}}$ it is of linear complexity $\mathcal{O}(N_{\mathbf{w}})$. A detailed complexity analysis for the latter case can be found in Appendix AG.

Note the differences between the Bayesian NLMS in Result 5.3 and the MAP-LMS algorithm in [82,83]. The latter one assumes the variable $\mathbf{z} = \mathbf{h} - \mathbf{w}_{k-1}$ follows a Gaussian distribution with zero mean, while the Bayesian NLMS algorithm assumes a Gaussian PDF of $\mathbf{h}$ with known mean $E_{\mathbf{h}}[\mathbf{h}]$ and covariance matrix $\mathbf{C}_{\mathbf{hh}}$. Another connection might be drawn to [84]. There, prior information is used on the model to incorporate systems with missing data and not as we do, on the parameter vector itself.

Example 5.3 (Estimation of an Impulse Response with the Bayesian NLMS Algorithm)

The goal in this example is to estimate an impulse response when prior knowledge is available. As prior knowledge, we choose $E_{\mathbf{h}}[\mathbf{h}] = \mathbf{0}^{N_\mathbf{w}}$ and $\mathbf{C}_{\mathbf{hh}} = 0.1\mathbf{I}^{N_\mathbf{w}\times N_\mathbf{w}}$, where $N_\mathbf{h} = N_\mathbf{w} = 5$. The number of measurements m is set to be 50 and a_k was set to $1/m$ for all k. The algorithms utilize $\epsilon = 10^{-3}$. In addition, a step-size reduction method is implemented that linearly decreases the step-size from 1 at $k = 1$ to $1/m$ at $k = m$. This step-size reduction method is utilized by the ordinary NLMS algorithm as well as by the Bayesian NLMS algorithm. The input samples $x[n]$ were generated from a complex proper Gaussian PDF with zero mean and unit variance. The noise samples were also randomly drawn from a complex proper Gaussian PDF with zero mean and variance σ_n^2. For the first investigation, σ_n^2 was set to 1. The convergence curves in terms of $E_{\mathbf{h},x,n}[||\mathbf{w}_k - \mathbf{h}||^2]$ are presented in Figure 5.5. These curves reveal that the Bayesian NLMS algorithm, since it incorporates prior knowledge about $\mathbf{h}$, outperforms the ordinary NLMS for all values of k.

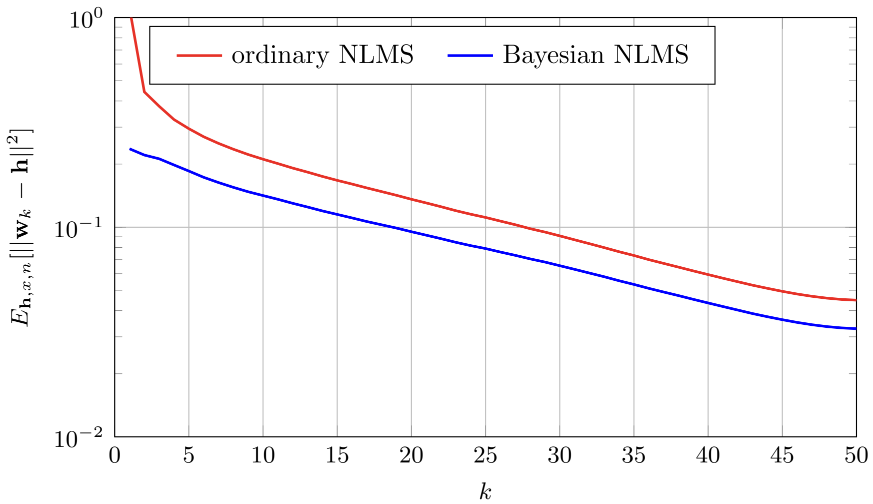

Figure 5.5: Convergence curves of the ordinary NLMS algorithm and the Bayesian NLMS algorithm. The complex-valued input and noise samples were both drawn from zero mean complex proper Gaussian distributions with unit variance. The true impulse responses were randomly drawn from a zero mean complex proper Gaussian distribution with covariance matrix $\mathbf{C}_{\mathbf{hh}} = 0.1\mathbf{I}^{5\times 5}$.

For the next investigation, σ_n^2 was varied between 10^{-2} and 10^2. For that, the performance in terms of $E_{\mathbf{h},x,n}[||\mathbf{w}_k - \mathbf{h}||^2]$ for $k = m = 50$ is visualized in Figure 5.6. In addition, the performance of the LS estimator and the LMMSE estimator are shown, which can be interpreted as the performance bounds for the RLS and sequential LMMSE algorithms, respectively. Figure 5.6 reveals that the Bayesian NLMS algorithm always performs better or equal to the ordinary NLMS algorithm. Also, the performance difference between the ordinary NLMS algorithm and the LS estimator is similar to the performance difference between the Bayesian NLMS algorithm and the LMMSE estimator.

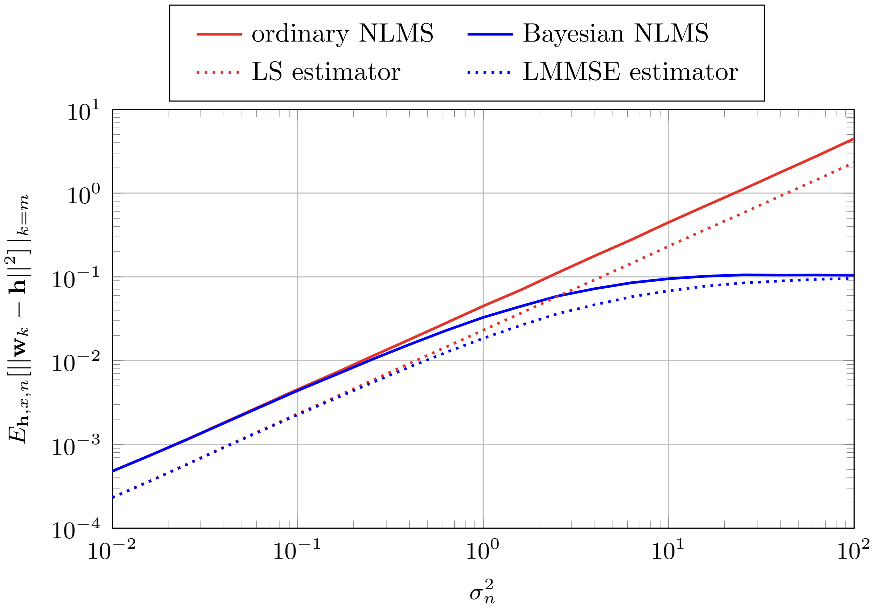

Figure 5.6: Convergence curves of the ordinary NLMS algorithm and the Bayesian NLMS algorithm. The complex-valued input and noise samples were both drawn from zero mean complex proper Gaussian distributions with unit variance. The true impulse responses were randomly drawn from a zero mean complex proper Gaussian distribution with covariance matrix $\mathbf{C}_{\mathbf{hh}} = 0.1\mathbf{I}^{5\times 5}$.

6 Conclusion

This thesis can be separated into three main parts.

The first main part considered classical estimation. There, methods that incorporate additional model knowledge into the estimation process in an optimal way were investigated. Four cases of additional model knowledge were considered.

The first case was the knowledge that the parameter vector of length n lies in a linear subspace of $\mathbb{C}^n$. It was proven that standard classical estimators such as the BLUE and the BWLUE can incorporate this additional knowledge in a straightforward manner. For the LS estimator on the other hand, it turned out that a constrained LS estimator is applicable, where the corresponding linear constraints were derived.

The second case of additional model knowledge considered in this thesis was the knowledge that the parameter vector fulfills additional linear constraints. In that case, the constrained LS estimator is available as a standard estimator but no corresponding extension for the BLUE and the BWLUE exists in the literature to the best of our knowledge. In this thesis, this gap was closed by proposing the constrained BLUE and the constrained BWLUE. It was shown that these novel estimators allow to increase the estimation accuracy compared to the BLUE and the BWLUE for the described scenario.

The third case was the knowledge that the parameter vector is real-valued while the measurements and the measurement noise are complex-valued. For that scenario, several widely linear classical estimators were proposed that incorporate this additional model knowledge in an optimal way. It was demonstrated that the resulting estimators outperform standard estimators as well as estimators that incorporate this additional model knowledge in an intuitive way.

The fourth case was the knowledge that the measurement matrix is subject to an unknown random error with known first and second order statistics. In this thesis, a novel algorithm was proposed that outperforms state-of-the-art algorithms significantly.

In the second main part, Bayesian estimators were investigated. Bayesian estimators consider the parameter vector to be random. This allows to include prior knowledge into the estimation process, in form of statistics of the parameter vector. Another difference between the classical and Bayesian approaches is the considered unbiased constraint. The unbiased constraint utilized by state-of-the-art Bayesian estimators is weaker than that utilized by unbiased classical estimators. Based on that, we investigated the so called component-wise conditionally unbiased (CWCU) constraints, which represent a trade-off between the stringent classical unbiased constraint and the usual weak Bayesian unbiased constraint. It was shown, that these unbiased constraints preserve the intuitive view of unbiasedness also in Bayesian scenarios, while allowing the incorporation of prior knowledge in many applications. Next, we focused on the class of so-called CWCU Bayesian estimators. We extended previous work on this type of estimator and extended the concept to widely linear estimators. The effects of these unbiased constraints, the relation to other Bayesian estimators and the ability to incorporate statistics about the unknown parameter vector were discussed.

Similar investigations as in the first main part of this thesis were performed in the third part in the context of adaptive filtering. Novel adaptive filters were derived that incorporate additional model knowledge that might be available in practice. The first sources of additional model knowledge incorporated by the derived adaptive filters was the knowledge that the optimal filter coefficients are real-valued whereas the input and desired signal are complex-valued. Again, the derived optimal adaptive filter algorithms significantly outperform state-of-the-art algorithms as well as intuitive algorithms in many applications. The second source of additional model knowledge concerns the task of system identification. For this case, a novel adaptive filter was proposed that incorporates prior knowledge about the impulse response of the unknown system. It was shown that the resulting algorithm features a reduced convergence time in the mean.

Appendices

A Commutation of the BWLUE Over Square Transformation Matrices

Reformulating (3.93) yields

$$\underline{\mathbf{x}} = \underline{\mathbf{B}}^{-1}(\underline{\boldsymbol{\alpha}} - \underline{\mathbf{c}}). \tag{A.1}$$

This expression inserted into the augmented linear model in (2.7) produces

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}\underline{\boldsymbol{\alpha}} - \underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}\underline{\mathbf{c}} + \underline{\mathbf{n}} \tag{A.2}$$

$$\underbrace{\underline{\mathbf{y}} + \underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}\underline{\mathbf{c}}}_{\underline{\tilde{\mathbf{y}}}} = \underbrace{\underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}}_{\underline{\tilde{\mathbf{H}}}}\,\underline{\boldsymbol{\alpha}} + \underline{\mathbf{n}} \tag{A.3}$$

$$\underline{\tilde{\mathbf{y}}} = \underline{\tilde{\mathbf{H}}}\,\underline{\boldsymbol{\alpha}} + \underline{\mathbf{n}}. \tag{A.4}$$

For this modified linear model, the BWLUE for $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}
\hat{\underline{\boldsymbol{\alpha}}}_{\mathrm{BW}} &= \left(\underline{\tilde{\mathbf{H}}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\tilde{\mathbf{H}}}\right)^{-1}\underline{\tilde{\mathbf{H}}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\tilde{\mathbf{y}}} && \text{(A.5)}\\
&= \left(\left(\underline{\mathbf{B}}^{-1}\right)^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}\right)^{-1}\left(\underline{\mathbf{B}}^{-1}\right)^H \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\left(\underline{\mathbf{y}} + \underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}\underline{\mathbf{c}}\right) && \text{(A.6)}\\
&= \underline{\mathbf{B}}\left(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\left(\underline{\mathbf{y}} + \underline{\mathbf{H}}\,\underline{\mathbf{B}}^{-1}\underline{\mathbf{c}}\right) && \text{(A.7)}\\
&= \underline{\mathbf{B}}\,\hat{\underline{\mathbf{x}}}_{\mathrm{BW}} + \underline{\mathbf{B}}\,\underline{\mathbf{B}}^{-1}\underline{\mathbf{c}} && \text{(A.8)}\\
&= \underline{\mathbf{B}}\,\hat{\underline{\mathbf{x}}}_{\mathrm{BW}} + \underline{\mathbf{c}}, && \text{(A.9)}
\end{aligned}$$

which concludes the proof. □

B Commutation of the BWLUE Over Rectangular Transformation Matrices

Consider the widely linear transformation in (3.93). The i^{th} row of this equation is given by $\alpha_i = \mathbf{b}_i^H \mathbf{x} + c_i$, where $\mathbf{b}_i^H$ is the i^{th} row of $\underline{\mathbf{B}}$ and where c_i is the i^{th} element of $\mathbf{c}$. We seek for a widely affine estimator of the form $\hat{\underline{\boldsymbol{\alpha}}} = \underline{\mathbf{E}}\underline{\mathbf{y}} + \underline{\mathbf{d}}$. Hence, the scalar $\hat{\alpha}_i$ is connected with the measurements via $\hat{\alpha}_i = \mathbf{e}_i^H \underline{\mathbf{y}} + d_i$, where $\mathbf{e}_i^H$ is the i^{th} row of the estimator matrix $\underline{\mathbf{E}}$ and where d_i is the i^{th} element of $\mathbf{d}$. Combining augmented form of the linear model in (3.1) with (3.93) leads to

$$\begin{aligned}
E_{\mathbf{y}}[\hat{\alpha}_i] &= E_{\mathbf{y}}\left[\mathbf{e}_i^H \underline{\mathbf{y}} + d_i\right] && \text{(B.1)}\\
&= E_{\mathbf{n}}\left[\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H \underline{\mathbf{n}} + d_i\right] && \text{(B.2)}\\
&= \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + d_i \stackrel{!}{=} \alpha_i, && \text{(B.3)}
\end{aligned}$$

or

$$\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + d_i \stackrel{!}{=} \mathbf{b}_i^H \underline{\mathbf{x}} + c_i. \tag{B.4}$$

To fulfill this for every $\mathbf{x}$ the conditions $\mathbf{e}_i^H \underline{\mathbf{H}} = \mathbf{b}_i^H$ and $d_i = c_i$ must hold. The cost function, which is the variance of $\hat{\alpha}_i$, follows as

$$J(\mathbf{e}_i) = E_{\mathbf{y}}\Big[(\hat{\alpha}_i - E_{\mathbf{y}}[\hat{\alpha}_i])(\hat{\alpha}_i - E_{\mathbf{y}}[\hat{\alpha}_i])^H\Big] \tag{B.5}$$

$$= E_{\mathbf{y}}\Big[\big(\mathbf{e}_i^H \underline{\mathbf{y}} + d_i - E_{\mathbf{y}}\big[\mathbf{e}_i^H \underline{\mathbf{y}} + d_i\big]\big)\big(\mathbf{e}_i^H \underline{\mathbf{y}} + d_i - E_{\mathbf{y}}\big[\mathbf{e}_i^H \underline{\mathbf{y}} + d_i\big]\big)^H\Big] \tag{B.6}$$

$$= E_{\mathbf{n}}\Big[\big(\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H \underline{\mathbf{n}} - \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}}\big)\big(\mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}} + \mathbf{e}_i^H \underline{\mathbf{n}} - \mathbf{e}_i^H \underline{\mathbf{H}}\,\underline{\mathbf{x}}\big)^H\Big] \tag{B.7}$$

$$= E_{\mathbf{n}}\Big[\big(\mathbf{e}_i^H \underline{\mathbf{n}}\big)\big(\mathbf{e}_i^H \underline{\mathbf{n}}\big)^H\Big] \tag{B.8}$$

$$= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i. \tag{B.9}$$

The vector $\mathbf{e}_i$ that minimizes this cost function and that produces unbiased estimates is the solution of the constrained optimization problem

$$\mathbf{e}_{\mathrm{BW},i} = \arg\min_{\mathbf{e}_i} \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{nn}} \mathbf{e}_i \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{H}} = \mathbf{b}_i^H, \tag{B.10}$$

where the index BW indicates the BWLUE. Solving this constrained optimization problem using the Lagrange multiplier method described in Section 2.4 leads to the BWLUE for α_i according to

$$\hat{\alpha}_{\mathrm{BW},i} = \mathbf{e}_{\mathrm{BW},i}^H \mathbf{y} + c_i \tag{B.11}$$

$$= \mathbf{b}_i^H \big(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\big)^{-1} \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{y}} + c_i. \tag{B.12}$$

Since $\mathbf{b}_i^H$ and c_i are the only terms that depend on the index i, the vector estimator immediately follows as

$$\hat{\underline{\boldsymbol{\alpha}}}_{\mathrm{BW}} = \underline{\mathbf{B}}\big(\underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{H}}\big)^{-1} \underline{\mathbf{H}}^H \underline{\mathbf{C}}_{\mathbf{nn}}^{-1} \underline{\mathbf{y}} + \underline{\mathbf{c}} \tag{B.13}$$

$$= \underline{\mathbf{B}}\,\hat{\underline{\mathbf{x}}}_{\mathrm{BW}} + \underline{\mathbf{c}}, \tag{B.14}$$

which proves the commutation of the BWLUE. □

C Proof that the Intuitive LS Estimator in (3.277) is Optimal if $\mathbf{H}^H\mathbf{H} = \alpha\mathbf{I}$

Consider the constrained LS estimator in (3.29). For the case of zero mean parameter vectors we choose $\mathbf{A} = \mathbf{1}^T$ and $\mathbf{b} = 0$, where $\mathbf{1}^T$ is a row vector of length $N_{\mathbf{x}}$ with all entries being 1. Then, the constrained LS estimator reads as

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{1}\mathbf{K}^{-1}\mathbf{1}^T\Big)\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{H}^H\mathbf{y} \tag{C.1}$$

$$= \Big(\mathbf{I} - \big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{1}\mathbf{K}^{-1}\mathbf{1}^T\Big)\hat{\mathbf{x}}_{\mathrm{OLS}}, \tag{C.2}$$

where

$$\mathbf{K} = \mathbf{1}^T\big(\mathbf{H}^H\mathbf{H}\big)^{-1}\mathbf{1}, \tag{C.3}$$

and where $\hat{\mathbf{x}}_{\text{OLS}}$ denotes the ordinary LS estimator. If the measurement matrix fulfills $\mathbf{H}^H\mathbf{H} = \alpha\mathbf{I}$, the estimator simplifies to

$$\hat{\mathbf{x}}_{\text{LS}} = \hat{\mathbf{x}}_{\text{OLS}} - \alpha^{-1}\mathbf{1}\left(\mathbf{1}^T\alpha^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^T\hat{\mathbf{x}}_{\text{OLS}} \tag{C.4}$$

$$= \hat{\mathbf{x}}_{\text{OLS}} - \mathbf{1}\underbrace{\left(\mathbf{1}^T\mathbf{1}\right)^{-1}}_{1/N_{\mathbf{x}}}\mathbf{1}^T\hat{\mathbf{x}}_{\text{OLS}} \tag{C.5}$$

$$= \hat{\mathbf{x}}_{\text{OLS}} - \mathbf{1}\underbrace{\frac{1}{N_{\mathbf{x}}}\mathbf{1}^T\hat{\mathbf{x}}_{\text{LS}}}_{\text{mean}(\hat{\mathbf{x}}_{\text{OLS}})} \tag{C.6}$$

$$= \hat{\mathbf{x}}_{\text{OLS}} - \text{mean}(\hat{\mathbf{x}}_{\text{OLS}})\mathbf{1}, \tag{C.7}$$

which corresponds to the intuitive estimator in (3.277). □

D Proof that (3.284) is Necessary and Sufficient to Make $\hat{x}_i$ Real-Valued

We now proof that the choice $\mathbf{f}_i^H = \mathbf{g}_i^T$ is necessary and sufficient to make $\hat{x}_i$ real-valued. With (3.281), it holds that

$$\text{Im}\{\hat{x}_i\} = \frac{1}{2}(\hat{x}_i - \hat{x}_i^*) \tag{D.1}$$

$$= \frac{1}{2}\left(\mathbf{f}_i^H\mathbf{y} + \mathbf{g}_i^H\mathbf{y}^* - \mathbf{f}_i^T\mathbf{y}^* - \mathbf{g}_i^T\mathbf{y}\right) \tag{D.2}$$

$$= \frac{1}{2}\left(\left(\mathbf{f}_i^H - \mathbf{g}_i^T\right)\mathbf{y} - \left(\mathbf{f}_i^T - \mathbf{g}_i^H\right)\mathbf{y}^*\right) \tag{D.3}$$

$$= \text{Im}\left\{\left(\mathbf{f}_i^H - \mathbf{g}_i^T\right)\mathbf{y}\right\}. \tag{D.4}$$

Since $\mathbf{y}$ in (D.4) is a complex-valued random variable, the only way to make (D.4) equal to zero for every possible realization of $\mathbf{y}$ is to enforce

$$\mathbf{f}_i^H - \mathbf{g}_i^T = \mathbf{0} \tag{D.5}$$

$$\mathbf{f}_i^H = \mathbf{g}_i^T. \tag{D.6}$$

Hence, $\mathbf{f}_i^H = \mathbf{g}_i^T$ is necessary to make $\hat{x}_i$ real-valued. The prove that $\mathbf{f}_i^H = \mathbf{g}_i^T$ is also sufficient to make $\hat{x}_i$ real-valued is obtained by inserting (D.6) into (D.4). □

E Variance and Pseudo-Variance of y_k

In the following, we make the approximation that $n_{A,k}$ has zero mean and variance $\sigma_{A,k}^2$ for $k = 1, \ldots, N_{\mathbf{y}} - 1$. The variance σ_k^2 of the k^{th} measurement y_k in Cartesian coordinates can be derived as

$$
\begin{aligned}
\sigma_k^2 &= E[(y_k - E[y_k])(y_k - E[y_k])^*] && \text{(E.1)}\\
&= E\Big[\big(A_k \mathrm{e}^{j\varphi_k}\mathrm{e}^{jn_{\varphi,k}} + n_{A,k}\mathrm{e}^{j\varphi_k}\mathrm{e}^{jn_{\varphi,k}} - \alpha_k A_k \mathrm{e}^{j\varphi_k}\big)\\
&\quad \times \big(A_k \mathrm{e}^{-j\varphi_k}\mathrm{e}^{-jn_{\varphi,k}} + n_{A,k}\mathrm{e}^{-j\varphi_k}\mathrm{e}^{-jn_{\varphi,k}} - \alpha_k A_k \mathrm{e}^{-j\varphi_k}\big)\Big] && \text{(E.2)}\\
&= \Big[A_k^2 + A_k n_{A,k} - \alpha_k A_k^2 \mathrm{e}^{jn_{\varphi,k}} + A_k n_{A,k} + n_{A,k}^2 - \alpha_k A_k n_{A,k}\mathrm{e}^{jn_{\varphi,k}}\\
&\quad - \alpha_k A_k^2 \mathrm{e}^{-jn_{\varphi,k}} - \alpha_k A_k n_{A,k}\mathrm{e}^{-jn_{\varphi,k}} + \alpha_k^2 A_k^2\Big] && \text{(E.3)}\\
&= A_k^2 - \alpha_k^2 A_k^2 + \sigma_{A,k}^2 - \alpha_k^2 A_k^2 + \alpha_k^2 A_k^2 && \text{(E.4)}\\
&= A_k^2(1 - \alpha_k^2) + \sigma_{A,k}^2. && \text{(E.5)}
\end{aligned}
$$

Similarly, the pseudo-variance $\widetilde{\sigma}_k^2$ of the k^{th} measurement y_k in Cartesian coordinates follows as

$$
\begin{aligned}
\widetilde{\sigma}_k^2 &= E[(y_k - E[y_k])(y_k - E[y_k])] && \text{(E.6)}\\
&= E\Big[\big(A_k \mathrm{e}^{j\varphi_k}\mathrm{e}^{jn_{\varphi,k}} + n_{A,k}\mathrm{e}^{j\varphi_k}\mathrm{e}^{jn_{\varphi,k}} - \alpha_k A_k \mathrm{e}^{j\varphi_k}\big)\\
&\quad \times \big(A_k \mathrm{e}^{j\varphi_k}\mathrm{e}^{jn_{\varphi,k}} + n_{A,k}\mathrm{e}^{j\varphi_k}\mathrm{e}^{jn_{\varphi,k}} - \alpha_k A_k \mathrm{e}^{j\varphi_k}\big)\Big] && \text{(E.7)}\\
&= E\Big[A_k^2 \mathrm{e}^{j2\varphi_k}\mathrm{e}^{j2n_{\varphi,k}} + 2A_k n_{A,k}\mathrm{e}^{j2\varphi_k}\mathrm{e}^{j2n_{\varphi,k}} - 2\alpha_k A_k^2 \mathrm{e}^{j2\varphi_k}\mathrm{e}^{jn_{\varphi,k}}\\
&\quad + n_{A,k}^2 \mathrm{e}^{j2\varphi_k}\mathrm{e}^{j2n_{\varphi,k}} - 2\alpha_k A_k n_{A,k}\mathrm{e}^{j2\varphi_k}\mathrm{e}^{jn_{\varphi,k}} + \alpha_k^2 A_k^2 \mathrm{e}^{j2\varphi_k}\Big] && \text{(E.8)}\\
&= A_k^2 \mathrm{e}^{j2\varphi_k}\underbrace{E[\mathrm{e}^{j2n_{\varphi,k}}]}_{\beta_k} - 2\alpha_k^2 A_k^2 \mathrm{e}^{j2\varphi_k} + \sigma_{A,k}^2 \mathrm{e}^{j2\varphi_k} E[\mathrm{e}^{j2n_{\varphi,k}}] + \alpha_k^2 A_k^2 \mathrm{e}^{j2\varphi_k} && \text{(E.9)}\\
&= A_k^2 \beta_k \mathrm{e}^{j2\varphi_k} - \alpha_k^2 A_k^2 \mathrm{e}^{j2\varphi_k} + \sigma_{A,k}^2 \beta_k \mathrm{e}^{j2\varphi_k} && \text{(E.10)}\\
&= \mathrm{e}^{j2\varphi_k}\big(\beta_k A_k^2 - \alpha_k^2 A_k^2 + \sigma_{A,k}^2 \beta_k\big). && \text{(E.11)}
\end{aligned}
$$

F Derivation of the Conditional MSE of the LMMSE Estimator

Consider the i^{th} estimate $\hat{x}_{\mathrm{L},i}$ in (4.13). For this estimator, the conditional MSE can be derived as

$$
\begin{aligned}
\operatorname{mse}(\hat{x}_{\mathrm{L},i}|\mathbf{x}) &= E_{\mathbf{y},\mathbf{x}}\big[|\hat{x}_{\mathrm{L},i} - x_i|^2|\mathbf{x}\big] && \text{(F.1)}\\
&= E_{\mathbf{y},\mathbf{x}}\big[|E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - x_i|^2|\mathbf{x}\big] && \text{(F.2)}\\
&= E_{\mathbf{x},\mathbf{n}}\big[| - (x_i - E_{x_i}[x_i]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{Hx} + \mathbf{n} - \mathbf{H}E_{\mathbf{x}}[\mathbf{x}])|^2|\mathbf{x}\big] && \text{(F.3)}\\
&= E_{\mathbf{x},\mathbf{n}}\big[| - (x_i - E_{x_i}[x_i]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{n}|^2|\mathbf{x}\big] && \text{(F.4)}\\
&= E_{\mathbf{x},\mathbf{n}}\big[\big(-(x_i - E_{x_i}[x_i]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{n}\big)\\
&\quad \times \big(-(x_i - E_{x_i}[x_i]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]) + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{n}\big)^H|\mathbf{x}\big] && \text{(F.5)}\\
&= E_{\mathbf{x},\mathbf{n}}\big[(x_i - E_{x_i}[x_i])(x_i - E_{x_i}[x_i])^* + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{nn}^H\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\\
&\quad + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^H\mathbf{H}^H\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\\
&\quad - (x_i - E_{x_i}[x_i])(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^H\mathbf{H}^H\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\\
&\quad - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])(x_i - E_{x_i}[x_i])^*|\mathbf{x}\big] && \text{(F.6)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\\
&\quad + \big((x_i - E_{x_i}[x_i]) - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big)(x_i - E_{x_i}[x_i])^*\\
&\quad - \big((x_i - E_{x_i}[x_i]) - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big)(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^H\mathbf{H}^H\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} && \text{(F.7)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\\
&\quad + \big(x_i - E_{x_i}[x_i] - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big)\\
&\quad \times \Big((x_i - E_{x_i}[x_i])^* - (\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])^H\mathbf{H}^H\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\Big) && \text{(F.8)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}\\
&\quad + \big(x_i - E_{x_i}[x_i] - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big)\\
&\quad \times \big(x_i - E_{x_i}[x_i] - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big)^* && \text{(F.9)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} + \big|x_i - E_{x_i}[x_i] - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])\big|^2. && \text{(F.10)}
\end{aligned}
$$

G Derivation of the Conditional Variance and the Conditional Bias of the LMMSE Estimator

Consider the i^{th} estimate $\hat{x}_{\mathrm{L},i}$ in (4.13). The proof for the conditional variance starts with the derivation of the conditional mean

$$
\begin{aligned}
E_{\mathbf{y}|\mathbf{x}}[\hat{x}_{\mathrm{L},i}|\mathbf{x}] &= E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\big(E_{\mathbf{y}|\mathbf{x}}[\mathbf{y}|\mathbf{x}] - E_{\mathbf{y}}[\mathbf{y}]\big) && \text{(G.1)}\\
&= E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{Hx} - \mathbf{H}E_{\mathbf{x}}[\mathbf{x}]). && \text{(G.2)}
\end{aligned}
$$

With that result, the conditional variance of $\hat{x}_i$ follows as

$$\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{L},i}|\mathbf{x}) &= E_{\mathbf{y}|\mathbf{x}}\left[\left|\hat{x}_{\mathrm{L},i} - E_{\mathbf{y}|\mathbf{x}}[\hat{x}_{\mathrm{L},i}|\mathbf{x}]\right|^2|\mathbf{x}\right] && \text{(G.3)}\\
&= E_{\mathbf{y}|\mathbf{x}}\Big[\Big|E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\\
&\qquad - E_{x_i}[x_i] - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{Hx} - \mathbf{H}E_{\mathbf{x}}[\mathbf{x}])\Big|^2|\mathbf{x}\Big] && \text{(G.4)}\\
&= E_{\mathbf{y}|\mathbf{x}}\left[\left|\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}] - \mathbf{Hx} + \mathbf{H}E_{\mathbf{x}}[\mathbf{x}])\right|^2|\mathbf{x}\right] && \text{(G.5)}\\
&= E_{\mathbf{n}}\left[\left|\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{n}\right|^2\right] && \text{(G.6)}\\
&= E_{\mathbf{n}}\left[\left(\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{n}\right)\left(\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{n}\right)^H\right] && \text{(G.7)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}E_{\mathbf{n}}\left[\mathbf{nn}^H\right]\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} && \text{(G.8)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{nn}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}. && \text{(G.9)}
\end{aligned}$$

The second proof concerns the conditional bias of $\hat{x}_{\mathrm{L},i}$, which directly follows from (G.2) as

$$\begin{aligned}
b(\hat{x}_{\mathrm{L},i}|\mathbf{x}) &= E_{\mathbf{y}|\mathbf{x}}[\hat{x}_{\mathrm{L},i} - x_i|\mathbf{x}] && \text{(G.10)}\\
&= E_{\mathbf{y}|\mathbf{x}}[\hat{x}_{\mathrm{L},i}|\mathbf{x}] - x_i && \text{(G.11)}\\
&= -x_i + E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{H}(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}]). && \text{(G.12)}
\end{aligned}$$

H Proof that the LMMSE Estimator Commutes Over Affine Transformations

In this appendix, we proof that the LMMSE estimator commutes over affine transformations. Let the transformation of the parameter vector be given by

$$\boldsymbol{\alpha} = \mathbf{Bx} + \mathbf{c}, \tag{H.1}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$, $\mathbf{B} \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$ and $\mathbf{c} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$. The LMMSE estimator for $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}_{\mathrm{L}} = E_{\boldsymbol{\alpha}}[\boldsymbol{\alpha}] + \mathbf{C}_{\boldsymbol{\alpha}\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{H.2}$$

Therein, $E_{\boldsymbol{\alpha}}[\boldsymbol{\alpha}]$ and $\mathbf{C}_{\boldsymbol{\alpha}\mathbf{y}}$ can be derived as

$$E_{\boldsymbol{\alpha}}[\boldsymbol{\alpha}] = \mathbf{B}E_{\mathbf{x}}[\mathbf{x}] + \mathbf{c} \tag{H.3}$$

and

$$\begin{aligned}
\mathbf{C}_{\boldsymbol{\alpha}\mathbf{y}} &= E_{\boldsymbol{\alpha},\mathbf{y}}\left[(\boldsymbol{\alpha} - E_{\boldsymbol{\alpha}}[\boldsymbol{\alpha}])(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])^H\right] && \text{(H.4)}\\
&= E_{\mathbf{x},\mathbf{y}}\left[(\mathbf{Bx} - \mathbf{B}E_{\mathbf{x}}[\mathbf{x}])(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])^H\right] && \text{(H.5)}\\
&= \mathbf{B}E_{\mathbf{x},\mathbf{y}}\left[(\mathbf{x} - E_{\mathbf{x}}[\mathbf{x}])(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])^H\right] && \text{(H.6)}\\
&= \mathbf{BC}_{\mathbf{xy}}, && \text{(H.7)}
\end{aligned}$$

respectively. Inserting (H.3) and (H.7) into the expression for $\hat{\boldsymbol{\alpha}}_{\mathrm{L}}$ in (H.2) leads to

$$\hat{\boldsymbol{\alpha}}_{\mathrm{L}} = \mathbf{B}E_{\mathbf{x}}[\mathbf{x}] + \mathbf{c} + \mathbf{B}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \tag{H.8}$$

$$= \mathbf{B}\hat{\mathbf{x}}_{\mathrm{L}} + \mathbf{c}, \tag{H.9}$$

which concludes the proof. □

I Proof that the WLMMSE Estimator Commutes Over Widely Affine Transformations

In this appendix, we proof that the WLMMSE estimator commutes over widely affine transformations. Let the transformation of the parameter vector be given by

$$\boldsymbol{\alpha} = \mathbf{B}_1\mathbf{x} + \mathbf{B}_2\mathbf{x}^* + \mathbf{c}, \tag{I.1}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$, $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{C}^{N_{\boldsymbol{\alpha}} \times N_{\mathbf{x}}}$ and $\mathbf{c} \in \mathbb{C}^{N_{\boldsymbol{\alpha}}}$. Eq. (I.1) can be brought into augmented form as

$$\underline{\boldsymbol{\alpha}} = \underline{\mathbf{B}}\,\underline{\mathbf{x}} + \underline{\mathbf{c}}, \tag{I.2}$$

where

$$\underline{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^* \end{bmatrix}, \quad \underline{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_2^* & \mathbf{B}_1^* \end{bmatrix}, \quad \underline{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \mathbf{c}^* \end{bmatrix}. \tag{I.3}$$

The WLMMSE estimator for $\boldsymbol{\alpha}$ in augmented notation is given by

$$\underline{\hat{\boldsymbol{\alpha}}}_{\mathrm{WL}} = E_{\boldsymbol{\alpha}}[\underline{\boldsymbol{\alpha}}] + \underline{\mathbf{C}}_{\boldsymbol{\alpha}\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\big). \tag{I.4}$$

Therein, $E_{\boldsymbol{\alpha}}[\underline{\boldsymbol{\alpha}}]$ and $\underline{\mathbf{C}}_{\boldsymbol{\alpha}\mathbf{y}}$ can be derived as

$$E_{\boldsymbol{\alpha}}[\underline{\boldsymbol{\alpha}}] = \underline{\mathbf{B}}E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{c}} \tag{I.5}$$

and

$$\underline{\mathbf{C}}_{\boldsymbol{\alpha}\mathbf{y}} = E_{\boldsymbol{\alpha},\mathbf{y}}\Big[(\underline{\boldsymbol{\alpha}} - E_{\boldsymbol{\alpha}}[\underline{\boldsymbol{\alpha}}])\big(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\big)^H\Big] \tag{I.6}$$

$$= E_{\mathbf{x},\mathbf{y}}\Big[(\underline{\mathbf{B}}\,\underline{\mathbf{x}} - \underline{\mathbf{B}}E_{\mathbf{x}}[\underline{\mathbf{x}}])\big(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\big)^H\Big] \tag{I.7}$$

$$= \underline{\mathbf{B}}E_{\mathbf{x},\mathbf{y}}\Big[(\underline{\mathbf{x}} - E_{\mathbf{x}}[\underline{\mathbf{x}}])\big(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\big)^H\Big] \tag{I.8}$$

$$= \underline{\mathbf{B}}\,\underline{\mathbf{C}}_{\mathbf{xy}}, \tag{I.9}$$

respectively. Inserting (I.5) and (I.9) into the expression for $\underline{\hat{\boldsymbol{\alpha}}}_{\mathrm{WL}}$ in (I.4) leads to

$$\underline{\hat{\boldsymbol{\alpha}}}_{\mathrm{WL}} = \underline{\mathbf{B}}E_{\mathbf{x}}[\underline{\mathbf{x}}] + \underline{\mathbf{c}} + \underline{\mathbf{B}}\underline{\mathbf{C}}_{\mathbf{xy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}\left[\underline{\mathbf{y}}\right]\big) \tag{I.10}$$

$$= \underline{\mathbf{B}}\underline{\hat{\mathbf{x}}}_{\mathrm{WL}} + \underline{\mathbf{c}}. \tag{I.11}$$

Considering only the upper half of (I.11) immediately leads to $\hat{\boldsymbol{\alpha}}_{\mathrm{WL}} = \mathbf{B}_1\hat{\mathbf{x}}_{\mathrm{WL}} + \mathbf{B}_2\hat{\mathbf{x}}_{\mathrm{WL}}^* + \mathbf{c}$. □

J Proof that the Linear Estimator Minimizing the BMSE Cost Function Subject to the Constraint in (4.71) Corresponds to the BLUE for an Underlying Linear Model

We focus on the i^{th} estimate $\hat{x}_i$ of the form

$$\hat{x}_i = \mathbf{e}_i^H \mathbf{y} + b_i \tag{J.1}$$

for this scalar estimator, the constraint in (4.71) reduces to

$$E_{\mathbf{y}|\mathbf{x}}[\hat{x}_i|\mathbf{x}] = x_i. \tag{J.2}$$

This constraint can be reformulated as

$$\begin{aligned}
E_{\mathbf{y}|\mathbf{x}}[\hat{x}_i|\mathbf{x}] =& E_{\mathbf{y}|\mathbf{x}}[\mathbf{e}_i^H \mathbf{y} + b_i|\mathbf{x}] && \text{(J.3)}\\
=& E_{\mathbf{y}|\mathbf{x}}[\mathbf{e}_i^H \mathbf{H}\mathbf{x} + \mathbf{e}_i^H \mathbf{n} + b_i|\mathbf{x}] && \text{(J.4)}\\
=& \mathbf{e}_i^H \mathbf{H}\mathbf{x} + b_i = x_i. && \text{(J.5)}
\end{aligned}$$

To fulfill this constraint for every $\mathbf{x}$, it must hold that $b_i = 0$ and $\mathbf{e}_i^H \mathbf{H} = \mathbf{u}_i^H$, where $\mathbf{u}_i^H$ is a row vector with a 1 at its i^{th} position and zero elsewhere. Incorporating these results into the BMSE cost function yields

$$\begin{aligned}
J(\mathbf{e}_i) =& E_{\mathbf{y},\mathbf{x}}\left[|x_i - \hat{x}_i|^2\right] && \text{(J.6)}\\
=& E_{\mathbf{y},\mathbf{x}}\left[\left(x_i - \mathbf{e}_i^H \mathbf{y} - b_i\right)\left(x_i - \mathbf{e}_i^H \mathbf{y} - b_i\right)^H\right] && \text{(J.7)}\\
=& E_{\mathbf{y},\mathbf{x}}\left[\left(x_i - \underbrace{\mathbf{e}_i^H \mathbf{H}\mathbf{x}}_{x_i} - \mathbf{e}_i^H \mathbf{n}\right)\left(x_i - \underbrace{\mathbf{e}_i^H \mathbf{H}\mathbf{x}}_{x_i} - \mathbf{e}_i^H \mathbf{n}\right)^H\right] && \text{(J.8)}\\
=& E_{\mathbf{n}}\left[\left(\mathbf{e}_i^H \mathbf{n}\right)\left(\mathbf{e}_i^H \mathbf{n}\right)^H\right] && \text{(J.9)}\\
=& \mathbf{e}_i^H \mathbf{C}_{\mathbf{n}\mathbf{n}} \mathbf{e}_i. && \text{(J.10)}
\end{aligned}$$

Inspecting the cost function in (J.10) as well as the constraint $\mathbf{e}_i^H \mathbf{H} = \mathbf{u}_i^H$ reveals that they correspond to the cost function and constraints utilized by the BLUE in Section 3.1. Hence, the optimization process will formally produce the BLUE, which is a classical estimator that does not utilize any prior knowledge. □

K Proof that $[\mathbf{D}]_{i,i} > 1$

The proof that $[\mathbf{D}]_{i,i} > 1$ is based on (4.152). Multiplying this equation with $\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}$ from the left and with $\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{y}x_i}$ produces

$$\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{y}\mathbf{y}|x_i}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{y}x_i} = \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{y}x_i} - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{y}x_i}. \tag{K.1}$$

Dividing this expression by $\sigma_{x_i}^2$ and utilizing the definition of $[\mathbf{D}]_{i,i}$ in (4.97) allows

$$\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yy}|x_i}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} = [\mathbf{D}]_{i,i}^{-1} - \left([\mathbf{D}]_{i,i}^{-1}\right)^2 \tag{K.2}$$

$$\frac{1}{\sigma_{x_i}^2}\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yy}|x_i}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} = [\mathbf{D}]_{i,i}^{-1}\left(1 - [\mathbf{D}]_{i,i}^{-1}\right). \tag{K.3}$$

The left hand side of (K.3) is positive and real-valued. The reason for this is that $\mathbf{C}_{\mathbf{yy}|x_i}$ is Hermitian and positive definite. Multiplying such a matrix with an arbitrary row vector from the left and with the conjugate transpose of this row vector from the right produces a real-valued scalar that is larger than zero. $[\mathbf{D}]_{i,i}^{-1}$ is also real-valued and larger than zero. Consequently, the expression in the brackets in (K.3) must be real-valued and larger than zero, too. This fact allows

$$1 - [\mathbf{D}]_{i,i}^{-1} > 0 \tag{K.4}$$

$$[\mathbf{D}]_{i,i}^{-1} < 1 \tag{K.5}$$

$$[\mathbf{D}]_{i,i} > 1, \tag{K.6}$$

concluding the proof. □

L Proof that (4.144) Corresponds to the CWCU LMMSE Estimator for Mutually Independent Parameters

For mutually independent parameters, $\mathbf{C}_{\mathbf{yy}}$ is given by

$$\begin{aligned}
\mathbf{C}_{\mathbf{yy}} &= E_{\mathbf{y},\mathbf{x}}\Big[(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])^H\Big] && \text{(L.1)}\\
&= E_{\mathbf{y},\mathbf{x}}\Big[\big(\mathbf{h}_i(x_i - E_{x_i}[x_i]) + \bar{\mathbf{H}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}\big)\\
&\qquad \times \big(\mathbf{h}_i(x_i - E_{x_i}[x_i]) + \bar{\mathbf{H}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}\big)^H\Big] && \text{(L.2)}\\
&= \mathbf{h}_i\sigma_{x_i}^2\mathbf{h}_i^H + \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_i^H + \mathbf{C}_{\mathbf{nn}} && \text{(L.3)}\\
&= \mathbf{h}_i\sigma_{x_i}^2\mathbf{h}_i^H + \mathbf{C}_i. && \text{(L.4)}
\end{aligned}$$

The inverse of $\mathbf{C}_{\mathbf{yy}}$ can be evaluated using Woodbury's matrix inversion lemma [54] and follows as

$$\mathbf{C}_{\mathbf{yy}}^{-1} = \mathbf{C}_i^{-1} - \mathbf{C}_i^{-1}\mathbf{h}_i\left(\frac{1}{\sigma_{x_i}^2} + \mathbf{h}_i^H\mathbf{C}_i^{-1}\mathbf{h}_i\right)^{-1}\mathbf{h}_i^H\mathbf{C}_i^{-1}. \tag{L.5}$$

With that, the first term of $\mathbf{e}_{\mathrm{CL},i}^H$ in (4.143) reads as

$$\frac{1}{\mathbf{h}_i^H \mathbf{C}_{\mathbf{yy}}^{-1} \mathbf{h}_i} = \left(\mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i - \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \left(\frac{1}{\sigma_{x_i}^2} + \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right)^{-1} \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right)^{-1} \tag{L.6}$$

$$= \left(\mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \left(1 - \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \left(\frac{1}{\sigma_{x_i}^2} + \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right)^{-1} \right) \right)^{-1} \tag{L.7}$$

$$= \frac{1}{\mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i} \left(1 - \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \left(\frac{1}{\sigma_{x_i}^2} + \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right)^{-1} \right)^{-1}. \tag{L.8}$$

The second term of $\mathbf{e}_{\mathrm{CL},i}^H$ in (4.143) reads as

$$\mathbf{h}_i^H \mathbf{C}_{\mathbf{yy}}^{-1} = \mathbf{h}_i^H \mathbf{C}_i^{-1} - \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \left(\frac{1}{\sigma_{x_i}^2} + \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right)^{-1} \mathbf{h}_i^H \mathbf{C}_i^{-1} \tag{L.9}$$

$$= \left(1 - \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \left(\frac{1}{\sigma_{x_i}^2} + \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right)^{-1} \right) \mathbf{h}_i^H \mathbf{C}_i^{-1}. \tag{L.10}$$

Inserting (L.8) and (L.10) into $\mathbf{e}_{\mathrm{CL},i}^H$ in (4.143) produces

$$\mathbf{e}_{\mathrm{CL},i}^H = \frac{1}{\mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i} \mathbf{h}_i^H \mathbf{C}_i^{-1}, \tag{L.11}$$

concluding the proof. $\square$

M Proof that (4.151) Holds for all Three Cases

The validity of (4.151) for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ is simply given by the properties of the Gaussian distribution [3]. We will now show that (4.151) also holds for the other two cases.

For the case when the linear model in (4.1) holds and when $\mathbf{x}$ is complex proper Gaussian, we have that

$$
\begin{aligned}
E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] &= E_{\bar{\mathbf{x}}_i,\mathbf{n}|x_i}\big[\mathbf{h}_i x_i + \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{n}|x_i\big] && \text{(M.1)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i] && \text{(M.2)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i\big(E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i])\big) && \text{(M.3)}\\
&= \big(\mathbf{h}_i + \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}\big)x_i + \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}E_{x_i}[x_i] && \text{(M.4)}\\
&= \underbrace{\big(\mathbf{h}_i\sigma_{x_i}^2 + \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big)}_{\mathbf{C}_{\mathbf{y}x_i}}(\sigma_{x_i}^2)^{-1}x_i + E_{\mathbf{y}}[\mathbf{y}] - \mathbf{h}_i E_{x_i}[x_i] - \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}E_{x_i}[x_i] && \text{(M.5)}\\
&= \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}x_i + E_{\mathbf{y}}[\mathbf{y}] - \underbrace{\big(\mathbf{h}_i\sigma_{x_i}^2 + \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big)}_{\mathbf{C}_{\mathbf{y}x_i}}(\sigma_{x_i}^2)^{-1}E_{x_i}[x_i] && \text{(M.6)}\\
&= \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}x_i + E_{\mathbf{y}}[\mathbf{y}] - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}E_{x_i}[x_i] && \text{(M.7)}\\
&= E_{\mathbf{y}}[\mathbf{y}] + \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]), && \text{(M.8)}
\end{aligned}
$$

where we utilized (4.116) and (4.122).

For the third case of mutually independent but otherwise arbitrary distributed elements of $\mathbf{x}$, we obtain with (4.136) that

$$
\begin{aligned}
E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] &= E_{\mathbf{y}|x_i}\big[\mathbf{h}_i x_i + \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{n}|x_i\big] && \text{(M.9)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && \text{(M.10)}\\
&= \mathbf{h}_i x_i + E_{\mathbf{y}}[\mathbf{y}] - \mathbf{h}_i E_{x_i}[x_i] && \text{(M.11)}\\
&= E_{\mathbf{y}}[\mathbf{y}] + \underbrace{\mathbf{h}_i\sigma_{x_i}^2}_{\mathbf{C}_{\mathbf{y}x_i}}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(M.12)}\\
&= E_{\mathbf{y}}[\mathbf{y}] + \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]), && \text{(M.13)}
\end{aligned}
$$

concluding the proof. □

N Proof that (4.152) Holds for all Three Cases

This appendix begins with the proof for the case of jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$.

For arbitrary jointly Gaussian vectors $\mathbf{z}$ and $\mathbf{w}$, the conditional covariance matrix is given by [3,85]

$$\mathbf{C}_{\mathbf{zz}|\mathbf{w}} = \mathbf{C}_{\mathbf{zz}} - \mathbf{C}_{\mathbf{zw}}\mathbf{C}_{\mathbf{ww}}^{-1}\mathbf{C}_{\mathbf{wz}}. \tag{N.1}$$

If $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian, x_i and $\mathbf{y}$ are jointly Gaussian, too. Adapting (N.1) to this case yields

$$\mathbf{C}_{\mathbf{yy}|x_i} = \mathbf{C}_{\mathbf{yy}} - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}. \tag{N.2}$$

For the second case where the linear model in (4.1) holds and where $\mathbf{x}$ is complex proper Gaussian, we first rewrite $\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]$ utilizing (4.122) and (4.151) as

$$\begin{aligned}
\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] &= \mathbf{y} - E_{\mathbf{y}}[\mathbf{y}] - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.3)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \mathbf{h}_i E_{x_i}[x_i] - \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.4)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \mathbf{h}_i E_{x_i}[x_i] - \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{h}_i\sigma_{x_i}^2(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) - \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.5)}\\
&= \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.6)}\\
&= \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \bar{\mathbf{H}}_i \underbrace{\left(E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i])\right)}_{E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]} && \text{(N.7)}\\
&= \bar{\mathbf{H}}_i\big(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]\big) + \mathbf{n}. && \text{(N.8)}
\end{aligned}$$

With that, the conditional covariance matrix can be derived as

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}\mathbf{y}|x_i} &= E_{\mathbf{y}|x_i}\Big[\big(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\big)\big(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\big)^H|x_i\Big] && \text{(N.9)}\\
&= E_{\mathbf{y}|x_i}\Big[\big(\bar{\mathbf{H}}_i\big(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]\big) + \mathbf{n}\big)\big(\bar{\mathbf{H}}_i\big(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]\big) + \mathbf{n}\big)^H|x_i\Big] && \text{(N.10)}\\
&= \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i|x_i}\bar{\mathbf{H}}_i^H + \mathbf{C}_{\mathbf{n}\mathbf{n}}. && \text{(N.11)}
\end{aligned}$$

Incorporating (4.122) and the fact that $\mathbf{x}$ is complex proper Gaussian yields

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}\mathbf{y}|x_i} &= \bar{\mathbf{H}}_i\big(\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i} - \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\big)\bar{\mathbf{H}}_i^H + \mathbf{C}_{\mathbf{n}\mathbf{n}} && \text{(N.12)}\\
&= \underbrace{\bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_i^H + \mathbf{h}_i\sigma_{x_i}^2\mathbf{h}_i^H + \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\mathbf{h}_i^H + \mathbf{h}_i\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_i^H + \mathbf{C}_{\mathbf{n}\mathbf{n}}}_{\mathbf{C}_{\mathbf{y}\mathbf{y}}}\\
&\quad - \mathbf{h}_i\sigma_{x_i}^2\mathbf{h}_i^H - \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\mathbf{h}_i^H - \mathbf{h}_i\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_i^H - \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_i^H && \text{(N.13)}\\
&= \mathbf{C}_{\mathbf{y}\mathbf{y}} - \underbrace{\big(\mathbf{h}_i\sigma_{x_i}^2 + \bar{\mathbf{H}}_i\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big)}_{\mathbf{C}_{\mathbf{y}x_i}}(\sigma_{x_i}^2)^{-1}\underbrace{\big(\sigma_{x_i}^2\mathbf{h}_i^H + \mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_i^H\big)}_{\mathbf{C}_{x_i\mathbf{y}}} && \text{(N.14)}\\
&= \mathbf{C}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}. && \text{(N.15)}
\end{aligned}$$

For the third case where the linear model in (4.1) holds and where the elements of $\mathbf{x}$ are uncorrelated, we first rewrite $\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]$ utilizing (4.136) and (4.151) as

$$\begin{aligned}
\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] &= \mathbf{y} - E_{\mathbf{y}}[\mathbf{y}] - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.16)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \mathbf{h}_i E_{x_i}[x_i] - \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.17)}\\
&= \mathbf{h}_i x_i + \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \mathbf{h}_i E_{x_i}[x_i] - \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{h}_i\sigma_{x_i}^2(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(N.18)}\\
&= \bar{\mathbf{H}}_i\bar{\mathbf{x}}_i + \mathbf{n} - \bar{\mathbf{H}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && \text{(N.19)}\\
&= \bar{\mathbf{H}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}. && \text{(N.20)}
\end{aligned}$$

With that and (4.136), the conditional covariance matrix can be derived as

$$\mathbf{C}_{\mathbf{yy}|x_i} = E_{\mathbf{y}|x_i}\left[\left(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\right)\left(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\right)^H |x_i\right] \tag{N.21}$$

$$= E_{\mathbf{y}|x_i}\left[\left(\bar{\mathbf{H}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}\right)\left(\bar{\mathbf{H}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}\right)^H |x_i\right] \tag{N.22}$$

$$= \bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i} \bar{\mathbf{H}}_i^H + \mathbf{C}_{\mathbf{nn}} \tag{N.23}$$

$$= \underbrace{\bar{\mathbf{H}}_i \mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i} \bar{\mathbf{H}}_i^H + \mathbf{h}_i \sigma_{x_i}^2 \mathbf{h}_i^H + \mathbf{C}_{\mathbf{nn}}}_{\mathbf{C}_{\mathbf{yy}}} - \mathbf{h}_i \sigma_{x_i}^2 \mathbf{h}_i^H \tag{N.24}$$

$$= \mathbf{C}_{\mathbf{yy}} - \underbrace{\mathbf{h}_i \sigma_{x_i}^2}_{\mathbf{C}_{\mathbf{y}x_i}} (\sigma_{x_i}^2)^{-1} \underbrace{\sigma_{x_i}^2 \mathbf{h}_i^H}_{\mathbf{C}_{x_i\mathbf{y}}} \tag{N.25}$$

$$= \mathbf{C}_{\mathbf{yy}} - \mathbf{C}_{\mathbf{y}x_i} (\sigma_{x_i}^2)^{-1} \mathbf{C}_{x_i\mathbf{y}}. \tag{N.26}$$

□

O Derivation of the Conditional Properties of the BLUE

Consider the BLUE for x_i in (3.47). For this estimator, the conditional mean follows as

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{B},i}|x_i] = \mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] \tag{O.1}$$

$$= \mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H} E_{\mathbf{x}|x_i}[\mathbf{x}|x_i] \tag{O.2}$$

$$= \mathbf{u}_i^H E_{\mathbf{x}|x_i}[\mathbf{x}|x_i] \tag{O.3}$$

$$= x_i. \tag{O.4}$$

With this result, the conditional bias immediately follows as

$$b(\hat{x}_{\mathrm{B},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{B},i}|x_i] - x_i = 0. \tag{O.5}$$

The conditional variance is given by

$$\mathrm{var}(\hat{x}_{\mathrm{B},i}|x_i) = E_{\mathbf{y}|x_i}\left[\left|\hat{x}_{\mathrm{B},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{B},i}|x_i]\right|^2 |x_i\right] \tag{O.6}$$

$$= E_{\mathbf{y}|x_i}\left[|\hat{x}_{\mathrm{B},i} - x_i|^2 |x_i\right] \tag{O.7}$$

$$= E_{\mathbf{y}|x_i}\left[\left|\mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{y} - x_i\right|^2 |x_i\right] \tag{O.8}$$

$$= E_{\mathbf{x},\mathbf{n}|x_i}\left[\left|\mathbf{u}_i^H \mathbf{x} + \mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{n} - x_i\right|^2 |x_i\right] \tag{O.9}$$

$$= E_{\mathbf{n}|x_i}\left[\left|\mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{n}\right|^2 |x_i\right] \tag{O.10}$$

$$= E_{\mathbf{n}}\left[\left|\mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{n}\right|^2\right] \tag{O.11}$$

$$= \mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{C}_{\mathbf{nn}} \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H} \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{u}_i \tag{O.12}$$

$$= \mathbf{u}_i^H \left(\mathbf{H}^H \mathbf{C}_{\mathbf{nn}}^{-1} \mathbf{H}\right)^{-1} \mathbf{u}_i. \tag{O.13}$$

Finally, the conditional MSE follows to

$$\mathrm{mse}(\hat{x}_{\mathrm{B},i}|x_i) = E_{\mathbf{y}|x_i}\Big[|\hat{x}_{\mathrm{B},i} - x_i|^2|x_i\Big] \tag{O.14}$$

$$= \mathrm{var}(\hat{x}_{\mathrm{B},i}|x_i) \tag{O.15}$$

$$= \mathbf{u}_i^H\big(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\big)^{-1}\mathbf{u}_i. \tag{O.16}$$

P Derivation of the Conditional Properties of the LMMSE Estimator

Consider the LMMSE for x_i in (4.13). For this estimator, the conditional mean follows as

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{L},i}|x_i] = E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\big(E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] - E_{\mathbf{y}}[\mathbf{y}]\big). \tag{P.1}$$

Inserting (4.151) produces

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{L},i}|x_i] = E_{x_i}[x_i] + \underbrace{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}}_{[\mathbf{D}]_{i,i}^{-1}}(x_i - E_{x_i}[x_i]) \tag{P.2}$$

$$= E_{x_i}[x_i] + [\mathbf{D}]_{i,i}^{-1}(x_i - E_{x_i}[x_i]) \tag{P.3}$$

$$= [\mathbf{D}]_{i,i}^{-1}x_i + (1 - [\mathbf{D}]_{i,i}^{-1})E_{x_i}[x_i]. \tag{P.4}$$

Now, the conditional bias can be derived as

$$b(\hat{x}_{\mathrm{L},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{L},i}|x_i] - x_i \tag{P.5}$$

$$= \Big([\mathbf{D}]_{i,i}^{-1} - 1\Big)x_i - \Big([\mathbf{D}]_{i,i}^{-1} - 1\Big)E_{x_i}[x_i] \tag{P.6}$$

$$= \Big([\mathbf{D}]_{i,i}^{-1} - 1\Big)(x_i - E_{x_i}[x_i]). \tag{P.7}$$

The conditional variance follows with (P.1) as

$$\mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i) = E_{\mathbf{y}|x_i}\Big[\big|\hat{x}_{\mathrm{L},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{L},i}|x_i]\big|^2|x_i\Big] \tag{P.8}$$

$$\begin{aligned} &= E_{\mathbf{y}|x_i}\Big[\big|E_{x_i}[x_i] + \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) - E_{x_i}[x_i] \\ &\qquad - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\big(E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] - E_{\mathbf{y}}[\mathbf{y}]\big)\big|^2|x_i\Big] \end{aligned} \tag{P.9}$$

$$= E_{\mathbf{y}|x_i}\Big[\big|\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\big(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\big)\big|^2|x_i\Big] \tag{P.10}$$

$$= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yy}|x_i}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}. \tag{P.11}$$

Assuming that one of the cases in Result 4.1 holds allows utilizing (4.152), resulting in

$$\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i) &= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\big(\mathbf{C}_{\mathbf{yy}} - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}\big)\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} && \text{(P.12)}\\
&= \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} - \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} && \text{(P.13)}\\
&= \sigma_{x_i}^2\left(\frac{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}}{\sigma_{x_i}^2} - \left(\frac{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}}{\sigma_{x_i}^2}\right)^2\right) && \text{(P.14)}\\
&= \sigma_{x_i}^2[\mathbf{D}]_{i,i}^{-1}\Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big). && \text{(P.15)}
\end{aligned}$$

With the conditional variance, the conditional bias and (4.98), the conditional MSE can be derived as

$$\begin{aligned}
\mathrm{mse}(\hat{x}_{\mathrm{L},i}|x_i) &= E_{\mathbf{y}|x_i}\Big[|\hat{x}_{\mathrm{L},i} - x_i|^2|x_i\Big] && \text{(P.16)}\\
&= \mathrm{var}(\hat{x}_{\mathrm{L},i}|x_i) + |b(\hat{x}_{\mathrm{L},i}|x_i)|^2 && \text{(P.17)}\\
&= \sigma_{x_i}^2[\mathbf{D}]_{i,i}^{-1}\Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big) + \Big|\Big([\mathbf{D}]_{i,i}^{-1} - 1\Big)(x_i - E_{x_i}[x_i])\Big|^2 && \text{(P.18)}\\
&= \sigma_{x_i}^2[\mathbf{D}]_{i,i}^{-1}\Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big) + \Big(1 - [\mathbf{D}]_{i,i}^{-1}\Big)^2|x_i - E_{x_i}[x_i]|^2. && \text{(P.19)}
\end{aligned}$$

Q Derivation of the Conditional Properties of the CWCU LMMSE Estimator

The CWCU LMMSE as given in (4.92) has the conditional mean

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i] = x_i \tag{Q.1}$$

and the conditional bias

$$b(\hat{x}_{\mathrm{CL},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i] - x_i = 0, \tag{Q.2}$$

since it fulfills the CWCU constraints. The conditional variance follows with (4.92) as

$$\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{CL},i}|x_i) &= E_{\mathbf{y}|x_i}\Big[\big|\hat{x}_{\mathrm{CL},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CL},i}|x_i]\big|^2|x_i\Big] && \text{(Q.3)}\\
&= E_{\mathbf{y}|x_i}\Big[\Big|E_{x_i}[x_i] + \mathbf{e}_{\mathrm{CL},i}^H(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}])\\
&\qquad - E_{x_i}[x_i] - \mathbf{e}_{\mathrm{CL},i}^H\big(E_{\mathbf{y}|x_i}[\mathbf{y}|x_i] - E_{\mathbf{y}}[\mathbf{y}]\big)\Big|^2|x_i\Big] && \text{(Q.4)}\\
&= E_{\mathbf{y}|x_i}\Big[\big|\mathbf{e}_{\mathrm{CL},i}^H\mathbf{y} - \mathbf{e}_{\mathrm{CL},i}^H E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\big|^2|x_i\Big] && \text{(Q.5)}\\
&= E_{\mathbf{y}|x_i}\Big[\big|\mathbf{e}_{\mathrm{CL},i}^H\big(\mathbf{y} - E_{\mathbf{y}|x_i}[\mathbf{y}|x_i]\big)\big|^2|x_i\Big] && \text{(Q.6)}\\
&= \mathbf{e}_{\mathrm{CL},i}^H\mathbf{C}_{\mathbf{yy}|x_i}\mathbf{e}_{\mathrm{CL},i}. && \text{(Q.7)}
\end{aligned}$$

Utilizing (4.91), (4.152) and (4.97), (Q.7) reads as

$$\operatorname{var}(\hat{x}_{\mathrm{CL},i}|x_i) = \mathbf{e}_{\mathrm{CL},i}^H\big(\mathbf{C}_{\mathbf{yy}} - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}\big)\mathbf{e}_{\mathrm{CL},i} \tag{Q.8}$$

$$= \left(\frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}}\right)^2 \mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\big(\mathbf{C}_{\mathbf{yy}} - \mathbf{C}_{\mathbf{y}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}}\big)\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i} \tag{Q.9}$$

$$= \frac{\big(\sigma_{x_i}^2\big)^2}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}} - \sigma_{x_i}^2 \tag{Q.10}$$

$$= \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). \tag{Q.11}$$

Finally, the conditional MSE can be derived as

$$\operatorname{mse}(\hat{x}_{\mathrm{CL},i}|x_i) = E_{\mathbf{y}|x_i}\Big[|\hat{x}_{\mathrm{CL},i} - x_i|^2|x_i\Big] \tag{Q.12}$$

$$= \operatorname{var}(\hat{x}_{\mathrm{CL},i}|x_i) \tag{Q.13}$$

$$= \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). \tag{Q.14}$$

R Proof that the CWCU LMMSE Estimator Coincides with the BLUE for diagonal $\mathbf{C}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{nn}}$ and $\mathbf{H}$

In this appendix, we prove that the CWCU LMMSE estimator coincides with the BLUE for the special case when $\mathbf{C}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{nn}}$ and $\mathbf{H}$ are all diagonal matrices, which implies that $N_{\mathbf{x}} = N_{\mathbf{y}} = N$. In this case, also $\mathbf{C}_{\mathbf{yy}} = \mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^H + \mathbf{C}_{\mathbf{nn}}$ is a diagonal matrix. We use the notation

$$\mathbf{H} = \operatorname{diag}\{[h_1, h_2, \ldots, h_N]\}, \tag{R.1}$$

$$\mathbf{C}_{\mathbf{xx}} = \operatorname{diag}\{[\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_N}^2]\}, \tag{R.2}$$

$$\mathbf{C}_{\mathbf{nn}} = \operatorname{diag}\{[\sigma_{n_1}^2, \sigma_{n_2}^2, \ldots, \sigma_{n_N}^2]\}. \tag{R.3}$$

Then, $\mathbf{C}_{x_i\mathbf{y}}$ becomes

$$\mathbf{C}_{x_i\mathbf{y}} = h_i^*\sigma_{x_i}^2\mathbf{u}_i^T, \tag{R.4}$$

where $\mathbf{u}_i^T$ is a length N row vector with a '1' at its i^{th} element and all zeros elsewhere. With that, we simplify the expression for the CWCU LMMSE estimator in (4.92) as

$$\hat{x}_{\mathrm{CL},i} = E_{x_i}[x_i] + \frac{\sigma_{x_i}^2}{\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{y}x_i}}\mathbf{C}_{x_i\mathbf{y}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \tag{R.5}$$

$$= E_{x_i}[x_i] + \sigma_{x_i}^2\frac{1}{h_i\sigma_{x_i}^2}\big(\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{u}_i\big)^{-1}\frac{1}{h_i^*\sigma_{x_i}^2}h_i^*\sigma_{x_i}^2\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) \tag{R.6}$$

$$= E_{x_i}[x_i] + \frac{1}{h_i}\big(\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{u}_i\big)^{-1}\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{R.7}$$

Consider the term $\big(\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{u}_i\big)^{-1}$ in (R.7). $\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{u}_i$ sorts out the i^{th} diagonal element of the diagonal matrix $\mathbf{C}_{\mathbf{yy}}^{-1}$. The inverse of this scalar corresponds to the i^{th} diagonal element of $\mathbf{C_{yy}}$. Hence, we can state that $\left(\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{u}_i\right)^{-1} = \mathbf{u}_i^T\mathbf{C_{yy}}\mathbf{u}_i$. This allows to modify (R.7) according to

$$\hat{x}_{\mathrm{CL},i} = E_{x_i}[x_i] + \frac{1}{h_i}\mathbf{u}_i^T\mathbf{C_{yy}}\mathbf{u}_i\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]). \tag{R.8}$$

The term $\mathbf{u}_i\mathbf{u}_i^T$ in (R.8) corresponds to a diagonal matrix (with only one diagonal element being non-zero). Diagonal matrices commute, hence, $\mathbf{C_{yy}}\mathbf{u}_i\mathbf{u}_i^T\mathbf{C}_{\mathbf{yy}}^{-1} = \mathbf{u}_i\mathbf{u}_i^T\mathbf{C_{yy}}\mathbf{C}_{\mathbf{yy}}^{-1} = \mathbf{u}_i\mathbf{u}_i^T$. Consequently, (R.8) corresponds to

$$\begin{aligned}
\hat{x}_{\mathrm{CL},i} &= E_{x_i}[x_i] + \frac{1}{h_i}\underbrace{\mathbf{u}_i^T\mathbf{u}_i}_{1}\mathbf{u}_i^T(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) && \text{(R.9)}\\
&= E_{x_i}[x_i] + \frac{1}{h_i}\mathbf{u}_i^T(\mathbf{y} - E_{\mathbf{y}}[\mathbf{y}]) && \text{(R.10)}\\
&= E_{x_i}[x_i] + \frac{1}{h_i}\mathbf{u}_i^T\mathbf{y} - \frac{1}{h_i}\underbrace{\mathbf{u}_i^T\mathbf{H}}_{h_i\mathbf{u}_i^T}E_{\mathbf{x}}[\mathbf{x}] && \text{(R.11)}\\
&= E_{x_i}[x_i] + \frac{1}{h_i}\mathbf{u}_i^T\mathbf{y} - \frac{1}{h_i}h_i\underbrace{\mathbf{u}_i^T E_{\mathbf{x}}[\mathbf{x}]}_{E_{x_i}[x_i]} && \text{(R.12)}\\
&= \frac{1}{h_i}\mathbf{u}_i^T\mathbf{y}. && \text{(R.13)}
\end{aligned}$$

This is the final result for the CWCU LMMSE estimator for the special case of diagonal $\mathbf{C_{xx}}$, $\mathbf{C_{nn}}$ and $\mathbf{H}$. We will now show that the BLUE formally yields the same expression since

$$\begin{aligned}
\hat{x}_{\mathrm{B},i} &= \mathbf{u}_i^T\left(\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} && \text{(R.14)}\\
&= \mathbf{u}_i^T\mathbf{H}^{-1}\mathbf{C_{nn}}\left(\mathbf{H}^H\right)^{-1}\mathbf{H}^H\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{y} && \text{(R.15)}\\
&= \underbrace{\mathbf{u}_i^T\mathbf{H}^{-1}}_{h_i^{-1}\mathbf{u}_i^T}\mathbf{y} && \text{(R.16)}\\
&= h_i^{-1}\mathbf{u}_i^T\mathbf{y}, && \text{(R.17)}
\end{aligned}$$

which corresponds to the expression of the CWCU LMMSE estimator in (R.13). □

S Derivation of Case 2 and 3 of Result 4.2

We start with case 2, assume a generalized complex Gaussian parameter vector $\mathbf{x}$, and begin the derivation of the i^{th} component $\hat{x}_i$ of the estimator. Recall the formulation of $\mathbf{y}$ in (4.113). We will now formulate a similar expression for $\underline{\mathbf{y}}$. With the notation

$$\begin{aligned}
\underline{\mathbf{x}}_i &= \begin{bmatrix} x_i \\ x_i^* \end{bmatrix} \in \mathbb{C}^2 & \underline{\bar{\mathbf{x}}}_i &= \begin{bmatrix} \bar{\mathbf{x}}_i \\ \bar{\mathbf{x}}_i^* \end{bmatrix} \in \mathbb{C}^{2N_{\mathbf{x}}-2}\\
\underline{\mathbf{H}}_i &= \begin{bmatrix} \mathbf{h}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_i^* \end{bmatrix} \in \mathbb{C}^{2N_{\mathbf{y}}\times 2}, & \underline{\bar{\mathbf{H}}}_i &= \begin{bmatrix} \bar{\mathbf{H}}_i & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_i^* \end{bmatrix} \in \mathbb{C}^{2N_{\mathbf{y}}\times(2N_{\mathbf{x}}-2)}
\end{aligned} \tag{S.1}$$

the augmented form of (4.113) becomes

$$\underline{\mathbf{y}} = \underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}}. \tag{S.2}$$

Incorporating (S.2) into the conditional mean of the estimator in (4.207) yields

$$\begin{aligned} E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] &= E_{\mathbf{y}|x_i}[\mathbf{e}_i^H\underline{\mathbf{y}} + b_i|x_i] \\ &= E_{\mathbf{n},\bar{\mathbf{x}}_i|x_i}[\mathbf{e}_i^H(\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}}) + b_i|x_i] \\ &= \mathbf{e}_i^H(\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]) + b_i. \end{aligned} \tag{S.3}$$

Because of the Gaussian assumption we have

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H\Big(\underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\big(E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] + \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])\big)\Big) + b_i. \tag{S.4}$$

By setting (S.4) equal to $x_i = \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{x}}_i$ one can see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if two conditions are fulfilled. The first condition is

$$\mathbf{e}_i^H\underline{\mathbf{H}}_i + \mathbf{e}_i^H\underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix}1 & 0\end{bmatrix} \tag{S.5}$$

$$\mathbf{e}_i^H\big(\underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i} + \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}\big) = \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i x_i}. \tag{S.6}$$

The expression in the brackets in (S.6) corresponds to $\underline{\mathbf{C}}_{\mathbf{y}x_i}$. This can be shown by

$$\underline{\mathbf{C}}_{\mathbf{y}x_i} = E_{\mathbf{y},\mathbf{x}}\Big[\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])^H\Big] \tag{S.7}$$

$$= E_{\mathbf{y},\mathbf{x}}\Big[\big(\underline{\mathbf{H}}_i(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) + \underline{\bar{\mathbf{H}}}_i(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i]) + \underline{\mathbf{n}}\big)(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])^H\Big] \tag{S.8}$$

$$= \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i} + \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}. \tag{S.9}$$

Hence, the first condition reads as

$$\mathbf{e}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i} = \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i x_i}. \tag{S.10}$$

The second condition follows from (S.4) as

$$b_i = -\mathbf{e}_i^H\underline{\bar{\mathbf{H}}}_i\big(E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] - \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}E_{x_i}[\underline{\mathbf{x}}_i]\big) \tag{S.11}$$

$$= -\mathbf{e}_i^H\underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] + \mathbf{e}_i^H\underbrace{\underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}}_{\underline{\mathbf{C}}_{\mathbf{y}x_i} - \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i}}\underline{\mathbf{C}}_{x_i x_i}^{-1}E_{x_i}[\underline{\mathbf{x}}_i] \tag{S.12}$$

$$= -\mathbf{e}_i^H\underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] - \mathbf{e}_i^H\underline{\mathbf{H}}_i E_{x_i}[\underline{\mathbf{x}}_i] + \underbrace{\mathbf{e}_i^H\underline{\mathbf{C}}_{\mathbf{y}x_i}}_{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i x_i}}\underline{\mathbf{C}}_{x_i x_i}^{-1}E_{x_i}[\underline{\mathbf{x}}_i]) \tag{S.13}$$

$$= -\mathbf{e}_i^H\underline{\mathbf{H}}E_{\mathbf{x}}[\underline{\mathbf{x}}] + \begin{bmatrix}1 & 0\end{bmatrix}E_{x_i}[\underline{\mathbf{x}}_i] \tag{S.14}$$

$$= E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}], \tag{S.15}$$

where (S.9) and (S.10) were utilized. Eq. (S.10) and (S.15) allow to simplify the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ according to

$$
\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \underline{\mathbf{y}} + b_i - x_i|^2] && \text{(S.16)}\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \begin{bmatrix}1 & 0\end{bmatrix}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])|^2] && \text{(S.17)}\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \mathbf{e}_i + \sigma_{x_i}^2 && \text{(S.18)}\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2 - \sigma_{x_i}^2 + \sigma_{x_i}^2 && \text{(S.19)}\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2. && \text{(S.20)}
\end{aligned}
$$

Combining the cost function in (S.20) and the constraint in (S.10) leads to the optimization problem

$$
\mathbf{e}_{\mathrm{CWL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2\right) \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix}1 & 0\end{bmatrix}. \tag{S.21}
$$

Note that this optimization problem equals the one for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ in (4.219) and solving it will lead to formally the same expression for the CWCU WLMMSE estimator. However, a significant difference is obtained. By making the assumption about an underlying linear model, the jointly Gaussian assumption of $\mathbf{x}$ and $\mathbf{y}$ can be significantly relaxed. In fact, only the parameter vector $\mathbf{x}$ is required to be Gaussian for obtaining (4.219). The PDF of the noise $\mathbf{n}$ can be arbitrary. The only requirements on the noise vector are $E_{\mathbf{n}}[\mathbf{n}] = \mathbf{0}$ while $\mathbf{n}$ and $\mathbf{x}$ need to be uncorrelated.

For mutually independent parameters (case 3 of Result 4.2) it is possible to further relax the prerequisites on $\mathbf{x}$. In this case (S.3) becomes

$$
E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H \underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + \mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] + b_i, \tag{S.22}
$$

since $E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]$ is no longer dependent on x_i. By setting (S.22) equal to $x_i = \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{x}}_i$ we see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$
\mathbf{e}_i^H \underline{\mathbf{H}}_i = \begin{bmatrix}1 & 0\end{bmatrix} \tag{S.23}
$$

and

$$
\begin{aligned}
b_i &= -\mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] && \text{(S.24)}\\
&= -\mathbf{e}_i^H \big(E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underline{\mathbf{H}}_i E_{x_i}[\underline{\mathbf{x}}_i]\big) && \text{(S.25)}\\
&= E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. && \text{(S.26)}
\end{aligned}
$$

Adapting (S.9) for the case of mutually independent parameters shows that in this case

$$
\underline{\mathbf{C}}_{\mathbf{y}x_i} = \underline{\mathbf{H}}_i \underline{\mathbf{C}}_{x_i x_i} \tag{S.27}
$$

holds. This result incorporated into (S.23) yields

$$
\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix}1 & 0\end{bmatrix}, \tag{S.28}
$$

which equals (S.10). Inserting in the BMSE cost function leads to the optimization problem

$$\mathbf{e}_{\mathrm{CWL},i} = \arg\min_{\mathbf{e}_i} \left(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{yy}} \mathbf{e}_i - \sigma_{x_i}^2\right) \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{S.29}$$

which again equals the one for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$ in (4.219). As a consequence, the expressions for the CWCU WLMMSE estimator are formally the same.

T Proof that (4.249) Holds for all Three Cases

For the case when the linear model in (4.1) holds and when $\mathbf{x}$ is Gaussian, it holds that

$$\begin{aligned}
E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] &= E_{\mathbf{y}|x_i}\left[\underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + \bar{\underline{\mathbf{H}}}_i \bar{\underline{\mathbf{x}}}_i + \underline{\mathbf{n}} | x_i\right] && \text{(T.1)}\\
&= \underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + \bar{\underline{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i|x_i}[\bar{\underline{\mathbf{x}}}_i|x_i] && \text{(T.2)}\\
&= \underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + \bar{\underline{\mathbf{H}}}_i \left(E_{\bar{\mathbf{x}}_i}[\bar{\underline{\mathbf{x}}}_i] + \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} (\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])\right) && \text{(T.3)}\\
&= \left(\underline{\mathbf{H}}_i + \bar{\underline{\mathbf{H}}}_i \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1}\right) \underline{\mathbf{x}}_i + \bar{\underline{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\underline{\mathbf{x}}}_i] - \bar{\underline{\mathbf{H}}}_i \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} E_{x_i}[\underline{\mathbf{x}}_i] && \text{(T.4)}\\
&= \underbrace{\left(\underline{\mathbf{H}}_i \underline{\mathbf{C}}_{x_i x_i} + \bar{\underline{\mathbf{H}}}_i \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}\right)}_{\underline{\mathbf{C}}_{\mathbf{y}x_i}} \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{x}}_i + E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underline{\mathbf{H}}_i E_{x_i}[\underline{\mathbf{x}}_i] \\
&\quad - \bar{\underline{\mathbf{H}}}_i \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} E_{x_i}[\underline{\mathbf{x}}_i] && \text{(T.5)}\\
&= \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{x}}_i + E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underbrace{\left(\underline{\mathbf{H}}_i \underline{\mathbf{C}}_{x_i x_i} + \bar{\underline{\mathbf{H}}}_i \underline{\mathbf{C}}_{\bar{\mathbf{x}}_i x_i}\right)}_{\underline{\mathbf{C}}_{\mathbf{y}x_i}} \underline{\mathbf{C}}_{x_i x_i}^{-1} E_{x_i}[\underline{\mathbf{x}}_i] && \text{(T.6)}\\
&= \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} \underline{\mathbf{x}}_i + E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} E_{x_i}[\underline{\mathbf{x}}_i] && \text{(T.7)}\\
&= E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} (\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]), && \text{(T.8)}
\end{aligned}$$

where we utilized (S.2) and (S.9).

For the third case of mutually independent but otherwise arbitrary distributed elements of $\mathbf{x}$, we obtain with (S.27)

$$\begin{aligned}
E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] &= E_{\mathbf{y}|x_i}\left[\underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + \bar{\underline{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\underline{\mathbf{x}}}_i] + \underline{\mathbf{n}} | x_i\right] && \text{(T.9)}\\
&= \underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + \bar{\underline{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && \text{(T.10)}\\
&= \underline{\mathbf{H}}_i \underline{\mathbf{x}}_i + E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underline{\mathbf{H}}_i E_{x_i}[\underline{\mathbf{x}}_i] && \text{(T.11)}\\
&= E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underbrace{\underline{\mathbf{H}}_i \underline{\mathbf{C}}_{x_i x_i}}_{\underline{\mathbf{C}}_{\mathbf{y}x_i}} \underline{\mathbf{C}}_{x_i x_i}^{-1} (\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) && \text{(T.12)}\\
&= E_{\mathbf{y}}[\underline{\mathbf{y}}] + \underline{\mathbf{C}}_{\mathbf{y}x_i} \underline{\mathbf{C}}_{x_i x_i}^{-1} (\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]), && \text{(T.13)}
\end{aligned}$$

concluding the proof. □

U Proof that (4.250) Holds for all Three Cases

This appendix begins with the proof for the case of jointly Gaussian $\mathbf{x}$ and $\mathbf{y}$.

For arbitrary jointly Gaussian vectors $\mathbf{z}$ and $\mathbf{w}$, the conditional covariance matrix is given by [3]

$$\underline{\mathbf{C}}_{\mathbf{zz}|\mathbf{w}} = \underline{\mathbf{C}}_{\mathbf{zz}} - \underline{\mathbf{C}}_{\mathbf{zw}}\underline{\mathbf{C}}_{\mathbf{ww}}^{-1}\underline{\mathbf{C}}_{\mathbf{wz}}. \tag{U.1}$$

If $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian, x_i and $\mathbf{y}$ are jointly Gaussian too. Adapting (U.1) to this case yields

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

For the second case where the linear model in (4.1) holds and where $\mathbf{x}$ is generalized complex Gaussian, we first rewrite $\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ utilizing (S.9) and (4.249) as

$$\begin{aligned}
\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] &= \underline{\mathbf{y}} - 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]) && \text{(U.3)}\\
&= \underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\mathbf{H}}_iE_{x_i}[\underline{\mathbf{x}}_i] - \underline{\bar{\mathbf{H}}}_iE_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i]\\
&\quad - \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]) && \text{(U.4)}\\
&= \underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\mathbf{H}}_iE_{x_i}[\underline{\mathbf{x}}_i] - \underline{\bar{\mathbf{H}}}_iE_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i]\\
&\quad - \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) - \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) && \text{(U.5)}\\
&= \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\bar{\mathbf{H}}}_iE_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] - \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) && \text{(U.6)}\\
&= \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\bar{\mathbf{H}}}_i\underbrace{\left(E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] - \underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])\right)}_{E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]} && \text{(U.7)}\\
&= \underline{\bar{\mathbf{H}}}_i\left(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]\right) + \underline{\mathbf{n}}. && \text{(U.8)}
\end{aligned}$$

With that, the conditional covariance matrix can be derived as

$$\begin{aligned}
\underline{\mathbf{C}}_{\mathbf{yy}|x_i} &= E_{\mathbf{y}|x_i}\left[\left(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\right)\left(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\right)^H|x_i\right] && \text{(U.9)}\\
&= E_{\mathbf{y}|x_i}\left[\left(\underline{\bar{\mathbf{H}}}_i\left(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]\right) + \underline{\mathbf{n}}\right)\left(\underline{\bar{\mathbf{H}}}_i\left(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\underline{\bar{\mathbf{x}}}_i|x_i]\right) + \underline{\mathbf{n}}\right)^H|x_i\right] && \text{(U.10)}\\
&= \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i|x_i}\underline{\bar{\mathbf{H}}}_i^H + \underline{\mathbf{C}}_{\mathbf{nn}}. && \text{(U.11)}
\end{aligned}$$

Incorporating (S.9) and the fact that $\mathbf{x}$ is generalized complex Gaussian yields

$$\begin{aligned}
\underline{\mathbf{C}}_{\mathbf{yy}|x_i} &= \underline{\bar{\mathbf{H}}}_i\left(\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i} - \underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\bar{\mathbf{x}}_i}\right)\underline{\bar{\mathbf{H}}}_i^H + \underline{\mathbf{C}}_{\mathbf{nn}} && \text{(U.12)}\\
&= \underbrace{\underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H + \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{H}}_i^H + \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{H}}_i^H + \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H + \underline{\mathbf{C}}_{\mathbf{nn}}}_{\underline{\mathbf{C}}_{\mathbf{yy}}}\\
&\quad - \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{H}}_i^H - \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{H}}_i^H - \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H - \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H && \text{(U.13)}\\
&= \underline{\mathbf{C}}_{\mathbf{yy}} - \underbrace{\left(\underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_ix_i} + \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_ix_i}\right)}_{\underline{\mathbf{C}}_{\mathbf{y}x_i}}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underbrace{\left(\underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{H}}_i^H + \underline{\mathbf{C}}_{x_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H\right)}_{\underline{\mathbf{C}}_{x_i\mathbf{y}}} && \text{(U.14)}\\
&= \underline{\mathbf{C}}_{\mathbf{yy}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}. && \text{(U.15)}
\end{aligned}$$

For the third case where the linear model in (4.1) holds and where the elements of $\mathbf{x}$ are uncorrelated, we first rewrite $\mathbf{y} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]$ utilizing (S.27) and (4.249) as

$$\begin{aligned}
\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] &= \underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}(x_i - E_{x_i}[\underline{\mathbf{x}}_i]) &(\text{U.16})\\
&= \underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\mathbf{H}}_i E_{x_i}[\underline{\mathbf{x}}_i] - \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i]\\
&\quad - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) &(\text{U.17})\\
&= \underline{\mathbf{H}}_i\underline{\mathbf{x}}_i + \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\mathbf{H}}_i E_{x_i}[\underline{\mathbf{x}}_i] - \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i]\\
&\quad - \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) &(\text{U.18})\\
&= \underline{\bar{\mathbf{H}}}_i\underline{\bar{\mathbf{x}}}_i + \underline{\mathbf{n}} - \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\underline{\bar{\mathbf{x}}}_i] &(\text{U.19})\\
&= \underline{\bar{\mathbf{H}}}_i(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \underline{\mathbf{n}}. &(\text{U.20})
\end{aligned}$$

With this result and (S.27), the conditional covariance matrix can be derived as

$$\begin{aligned}
\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i} &= E_{\mathbf{y}|x_i}\left[\left(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\right)\left(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\right)^H|x_i\right] &(\text{U.21})\\
&= E_{\mathbf{y}|x_i}\left[\left(\underline{\bar{\mathbf{H}}}_i(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \underline{\mathbf{n}}\right)\left(\underline{\bar{\mathbf{H}}}_i(\underline{\bar{\mathbf{x}}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \underline{\mathbf{n}}\right)^H|x_i\right] &(\text{U.22})\\
&= \underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H + \underline{\mathbf{C}}_{\mathbf{n}\mathbf{n}} &(\text{U.23})\\
&= \underbrace{\underline{\bar{\mathbf{H}}}_i\underline{\mathbf{C}}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\underline{\bar{\mathbf{H}}}_i^H + \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i}\underline{\mathbf{H}}_i^H + \underline{\mathbf{C}}_{\mathbf{n}\mathbf{n}}}_{\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}} - \underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i}\underline{\mathbf{H}}_i^H &(\text{U.24})\\
&= \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \underbrace{\underline{\mathbf{H}}_i\underline{\mathbf{C}}_{x_i x_i}}_{\underline{\mathbf{C}}_{\mathbf{y}x_i}}\underline{\mathbf{C}}_{x_i x_i}^{-1}\underbrace{\underline{\mathbf{C}}_{x_i x_i}\underline{\mathbf{H}}_i^H}_{\underline{\mathbf{C}}_{x_i\mathbf{y}}} &(\text{U.25})\\
&= \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_i x_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}. &(\text{U.26})
\end{aligned}$$

□

V Derivation of the Conditional Properties of the BWLUE

Consider the BWLUE for x_i in (3.62). For this estimator, the conditional mean follows as

$$\begin{aligned}
E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{BW},i}|x_i] &= \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{n}\mathbf{n}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{n}\mathbf{n}}^{-1}E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] &(\text{V.1})\\
&= \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{n}\mathbf{n}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{n}\mathbf{n}}^{-1}\underline{\mathbf{H}}E_{\mathbf{x}|x_i}[\underline{\mathbf{x}}|x_i] &(\text{V.2})\\
&= \mathbf{u}_i^H E_{\mathbf{x}|x_i}[\underline{\mathbf{x}}|x_i] &(\text{V.3})\\
&= x_i. &(\text{V.4})
\end{aligned}$$

With this result, the conditional bias becomes

$$b(\hat{x}_{\mathrm{BW},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{BW},i}|x_i] - x_i = 0. \tag{V.5}$$

The conditional variance is given by

$$
\begin{aligned}
\operatorname{var}(\hat{x}_{\mathrm{BW},i}|x_i) &= E_{\mathbf{y}|x_i}\left[\left|\hat{x}_{\mathrm{BW},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{BW},i}|x_i]\right|^2 |x_i\right] && \text{(V.6)}\\
&= E_{\mathbf{y}|x_i}\left[|\hat{x}_{\mathrm{BW},i} - x_i|^2 |x_i\right] && \text{(V.7)}\\
&= E_{\mathbf{y}|x_i}\left[\left|\mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} - x_i\right|^2 |x_i\right] && \text{(V.8)}\\
&= E_{\mathbf{x},\mathbf{n}|x_i}\left[\left|\mathbf{u}_i^H\underline{\mathbf{x}} + \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{n}} - x_i\right|^2 |x_i\right] && \text{(V.9)}\\
&= E_{\mathbf{n}|x_i}\left[\left|\mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{n}}\right|^2 |x_i\right] && \text{(V.10)}\\
&= \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{C}}_{\mathbf{nn}}\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\mathbf{u}_i && \text{(V.11)}\\
&= \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\mathbf{u}_i. && \text{(V.12)}
\end{aligned}
$$

Finally, the conditional MSE can be derived as

$$
\begin{aligned}
\operatorname{mse}(\hat{x}_{\mathrm{BW},i}|x_i) &= E_{\mathbf{y}|x_i}\left[|\hat{x}_{\mathrm{BW},i} - x_i|^2 |x_i\right] && \text{(V.13)}\\
&= \operatorname{var}(\hat{x}_{\mathrm{BW},i}|x_i) && \text{(V.14)}\\
&= \mathbf{u}_i^H\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\mathbf{u}_i. && \text{(V.15)}
\end{aligned}
$$

W Derivation of the Conditional Properties of the WLMMSE Estimator

Consider the WLMMSE for x_i in (4.49). For this estimator, the conditional mean follows as

$$
E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] = E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right). \tag{W.1}
$$

Inserting (4.249) produces

$$
\begin{aligned}
E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] &= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underbrace{\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}}_{\underline{\mathbf{D}}_i^{-1}}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) && \text{(W.2)}\\
&= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) && \text{(W.3)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix}E_{x_i}[\underline{\mathbf{x}}_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) && \text{(W.4)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix}\left(\underline{\mathbf{D}}_i^{-1}\underline{\mathbf{x}}_i + (\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1})E_{x_i}[\underline{\mathbf{x}}_i]\right). && \text{(W.5)}
\end{aligned}
$$

Now, the conditional bias can be derived as

$$
\begin{aligned}
b(\hat{x}_{\mathrm{WL},i}|x_i) &= E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] - x_i && \text{(W.6)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} E_{x_i}[\underline{\mathbf{x}}_i] + \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{D}}_i^{-1}(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]) - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{x}}_i && \text{(W.7)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)\underline{\mathbf{x}}_i - \begin{bmatrix}1 & 0\end{bmatrix} \big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big) E_{x_i}[\underline{\mathbf{x}}_i] && \text{(W.8)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]). && \text{(W.9)}
\end{aligned}
$$

The conditional variance follows with (W.1) as

$$
\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) &= E_{\mathbf{y}|x_i}\Big[\big|\hat{x}_{\mathrm{WL},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i]\big|^2 |x_i\Big] && \text{(W.10)}\\
&= E_{\mathbf{y}|x_i}\Big[\Big|E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - E_{x_i}[x_i] \\
&\qquad - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\Big|^2 |x_i\Big] && \text{(W.11)}\\
&= E_{\mathbf{y}|x_i}\Big[\Big|\begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\big)\Big|^2 |x_i\Big] && \text{(W.12)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{yy}|x_i}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}. && \text{(W.13)}
\end{aligned}
$$

Utilizing (4.250) allows

$$
\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) &= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\big(\underline{\mathbf{C}}_{\mathbf{yy}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\big)\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix} && \text{(W.14)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix} \\
&\quad - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\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} && \text{(W.15)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \underbrace{\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}}_{\underline{\mathbf{D}}_i^{-1}}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} \\
&\quad - \begin{bmatrix}1 & 0\end{bmatrix} \underbrace{\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}}_{\underline{\mathbf{D}}_i^{-1}}\underbrace{\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}}_{\underline{\mathbf{D}}_i^{-1}}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} && \text{(W.16)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{D}}_i^{-1}\underline{\mathbf{D}}_i^{-1}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} && \text{(W.17)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix}. && \text{(W.18)}
\end{aligned}
$$

With the conditional variance and the conditional bias, the conditional MSE can be derived as

$$
\begin{aligned}
\operatorname{mse}(\hat{x}_{\mathrm{WL},i}|x_i) &= E_{\mathbf{y}|x_i}\Big[|\hat{x}_{\mathrm{WL},i} - x_i|^2|x_i\Big] && \text{(W.19)}\\
&= \operatorname{var}(\hat{x}_{\mathrm{WL},i}|x_i) + |b(\hat{x}_{\mathrm{WL},i}|x_i)|^2 && \text{(W.20)}\\
&= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i^{-1}\big(\mathbf{I}^{2\times 2} - \underline{\mathbf{D}}_i^{-1}\big)\underline{\mathbf{C}}_{x_i x_i}\begin{bmatrix}1\\0\end{bmatrix}\\
&\quad + \Big|\begin{bmatrix}1 & 0\end{bmatrix}\big(\underline{\mathbf{D}}_i^{-1} - \mathbf{I}^{2\times 2}\big)(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])\Big|^2. && \text{(W.21)}
\end{aligned}
$$

X Derivation of the Conditional Properties of the CWCU WLMMSE Estimator

The CWCU WLMMSE estimator from (4.228) has the conditional mean

$$
E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] = x_i \tag{X.1}
$$

and the conditional bias

$$
b(\hat{x}_{\mathrm{CWL},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] - x_i = 0 \tag{X.2}
$$

since it fulfills the CWCU constraints. The conditional variance follows with (4.228) as

$$
\begin{aligned}
\operatorname{var}(\hat{x}_{\mathrm{CWL},i}|x_i) &= E_{\mathbf{y}|x_i}\Big[\big|\hat{x}_{\mathrm{CWL},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i]\big|^2|x_i\Big] && \text{(X.3)}\\
&= E_{\mathbf{y}|x_i}\Big[\Big|E_{x_i}[x_i] + \mathbf{e}_{\mathrm{CWL},i}^H\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\\
&\quad - E_{x_i}[x_i] - \mathbf{e}_{\mathrm{CWL},i}^H\big(E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\Big|^2|x_i\Big] && \text{(X.4)}\\
&= E_{\mathbf{y}|x_i}\Big[\big|\mathbf{e}_{\mathrm{CWL},i}^H\underline{\mathbf{y}} - \mathbf{e}_{\mathrm{CWL},i}^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\big|^2|x_i\Big] && \text{(X.5)}\\
&= E_{\mathbf{y}|x_i}\Big[\big|\mathbf{e}_{\mathrm{CWL},i}^H\big(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\big)\big|^2|x_i\Big] && \text{(X.6)}\\
&= \mathbf{e}_{\mathrm{CWL},i}^H\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i}\mathbf{e}_{\mathrm{CWL},i}. && \text{(X.7)}
\end{aligned}
$$

Utilizing (4.226), (4.227) and (4.250), (X.7) reads as

$$\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) = \mathbf{e}_{\mathrm{CWL},i}^H\big(\underline{\mathbf{C}}_{\mathbf{yy}} - \underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\big)\mathbf{e}_{\mathrm{CWL},i} \tag{X.8}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{D}}_i^H\begin{bmatrix}1\\0\end{bmatrix}$$

$$- \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{C}}_{x_ix_i}^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\underline{\mathbf{D}}_i^H\begin{bmatrix}1\\0\end{bmatrix} \tag{X.9}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} \tag{X.10}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} - \sigma_{x_i}^2. \tag{X.11}$$

Finally, the conditional MSE can be derived as

$$\mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i) = E_{\mathbf{y}|x_i}\left[|\hat{x}_{\mathrm{CWL},i} - x_i|^2|x_i\right] \tag{X.12}$$

$$= \mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) \tag{X.13}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{D}}_i\underline{\mathbf{C}}_{x_ix_i}\begin{bmatrix}1\\0\end{bmatrix} - \sigma_{x_i}^2. \tag{X.14}$$

Y Proof that the CWCU WLMMSE Estimator Coincides with the BWLUE for Diagonal $\mathbf{C_{xx}}$, $\widetilde{\mathbf{C}}_{\mathbf{xx}}$, $\mathbf{C_{nn}}$, $\widetilde{\mathbf{C}}_{\mathbf{nn}}$ and $\mathbf{H}$

In this appendix, we prove that the CWCU WLMMSE estimator coincides with the BWLUE for the special case when $\mathbf{C_{xx}}$, $\widetilde{\mathbf{C}}_{\mathbf{xx}}$, $\mathbf{C_{nn}}$, $\widetilde{\mathbf{C}}_{\mathbf{nn}}$ and $\mathbf{H}$ are all diagonal matrices. In addition to the notation in (R.1)–(R.3), we utilize

$$\widetilde{\mathbf{C}}_{\mathbf{xx}} = \mathrm{diag}\{[\widetilde{\sigma}_{x_1}^2, \widetilde{\sigma}_{x_2}^2, \dots, \widetilde{\sigma}_{x_N}^2]\} \tag{Y.1}$$

and

$$\widetilde{\mathbf{C}}_{\mathbf{nn}} = \mathrm{diag}\{[\widetilde{\sigma}_{n_1}^2, \widetilde{\sigma}_{n_2}^2, \dots, \widetilde{\sigma}_{n_N}^2]\}. \tag{Y.2}$$

With $\mathbf{h}_i$ denoting the i^{th} column of $\mathbf{H}$ and h_i denoting the i^{th} diagonal element of $\mathbf{H}$, we define

$$\underline{\mathbf{H}}_i = \begin{bmatrix}\mathbf{h}_i & \mathbf{0}^N\\ \mathbf{0}^N & \mathbf{h}_i^*\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{u}_i & \mathbf{0}^N\\ \mathbf{0}^N & \mathbf{u}_i\end{bmatrix}}_{\underline{\mathbf{U}}_i}\underbrace{\begin{bmatrix}h_i & 0\\ 0 & h_i^*\end{bmatrix}}_{\widetilde{\underline{\mathbf{H}}}_i} \tag{Y.3}$$

$$= \underline{\mathbf{U}}_i\widetilde{\underline{\mathbf{H}}}_i, \tag{Y.4}$$

where $\mathbf{u}_i$ is a column vector with a '1' at its i^{th} element and all zeros elsewhere. Now, $\underline{\mathbf{C}}_{x_i\mathbf{y}}$ becomes

$$\underline{\mathbf{C}}_{x_i\mathbf{y}} = \underline{\mathbf{C}}_{x_ix_i}\underline{\mathbf{H}}_i^H = \underline{\mathbf{C}}_{x_ix_i}\widetilde{\underline{\mathbf{H}}}_i^H\underline{\mathbf{U}}_i^T. \tag{Y.5}$$

With that, we simplify the expression for the CWCU WLMMSE estimator in (4.228) as

$$\hat{x}_{\mathrm{CWL},i} = E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\right)^{-1}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{Y.6}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_ix_i}\left(\underline{\mathbf{C}}_{x_ix_i}\underline{\widetilde{\mathbf{H}}}_i^H\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{U}}_i\underline{\widetilde{\mathbf{H}}}_i\underline{\mathbf{C}}_{x_ix_i}\right)^{-1}\underline{\mathbf{C}}_{x_ix_i}\underline{\widetilde{\mathbf{H}}}_i^H\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{Y.7}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\left(\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{U}}_i\right)^{-1}\left(\underline{\widetilde{\mathbf{H}}}_i^H\right)^{-1}\underline{\widetilde{\mathbf{H}}}_i^H\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{Y.8}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\left(\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{U}}_i\right)^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right). \tag{Y.9}$$

Similar to Appendix R, one can show that for the considered special case it holds that $\left(\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\underline{\mathbf{U}}_i\right)^{-1} = \underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}\underline{\mathbf{U}}_i$ and that $\underline{\mathbf{C}}_{\mathbf{yy}}\underline{\mathbf{U}}_i\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}^{-1} = \underline{\mathbf{U}}_i\underline{\mathbf{U}}_i^T\underline{\mathbf{C}}_{\mathbf{yy}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1} = \underline{\mathbf{U}}_i\underline{\mathbf{U}}_i^T$. This allows to modify (Y.9) according to

$$\hat{x}_{\mathrm{CWL},i} = E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underbrace{\underline{\mathbf{U}}_i^T\underline{\mathbf{U}}_i}_{\mathbf{I}^{2\times 2}}\underline{\mathbf{U}}_i^T\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{Y.10}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\left(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right) \tag{Y.11}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{y}} - \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underbrace{\underline{\mathbf{U}}_i^T\underline{\mathbf{H}}}_{\underline{\mathbf{H}}_i^T}E_{\mathbf{x}}[\underline{\mathbf{x}}] \tag{Y.12}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{y}} - \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{H}}_i^T E_{\mathbf{x}}[\underline{\mathbf{x}}] \tag{Y.13}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{y}} - \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underbrace{\underline{\widetilde{\mathbf{H}}}_i^T}_{\underline{\widetilde{\mathbf{H}}}_i}\underbrace{\underline{\mathbf{U}}_i^T E_{\mathbf{x}}[\underline{\mathbf{x}}]}_{E_{x_i}[\underline{\mathbf{x}}_i]} \tag{Y.14}$$

$$= E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{y}} - \underbrace{\begin{bmatrix}1 & 0\end{bmatrix}E_{x_i}[\underline{\mathbf{x}}_i]}_{E_{x_i}[x_i]} \tag{Y.15}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{y}}. \tag{Y.16}$$

This is the final result for the CWCU WLMMSE estimator for the special case of diagonal $\mathbf{C}_{\mathbf{xx}}$, $\widetilde{\mathbf{C}}_{\mathbf{xx}}$, $\mathbf{C}_{\mathbf{nn}}$, $\widetilde{\mathbf{C}}_{\mathbf{nn}}$ and $\mathbf{H}$. We will now show that the BWLUE formally yields the same expression. This can be shown as

$$\hat{x}_{\mathrm{WB},i} = \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{U}}_i^T\left(\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{H}}\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} \tag{Y.17}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{U}}_i^T\underline{\mathbf{H}}^{-1}\underline{\mathbf{C}}_{\mathbf{nn}}\left(\underline{\mathbf{H}}^H\right)^{-1}\underline{\mathbf{H}}^H\underline{\mathbf{C}}_{\mathbf{nn}}^{-1}\underline{\mathbf{y}} \tag{Y.18}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underbrace{\underline{\mathbf{U}}_i^T\underline{\mathbf{H}}^{-1}}_{=\left(\underline{\mathbf{H}}_i^T\right)^{-1}=\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T}\underline{\mathbf{y}} \tag{Y.19}$$

$$= \begin{bmatrix}1 & 0\end{bmatrix}\underline{\widetilde{\mathbf{H}}}_i^{-1}\underline{\mathbf{U}}_i^T\underline{\mathbf{y}}, \tag{Y.20}$$

which corresponds to the expression of the CWCU WLMMSE estimator in (Y.16). □

Z Derivation of Case 2 and 3 of Result 4.3

We start with case 2, which assumes a real-valued Gaussian parameter vector $\mathbf{x}$, and begin the derivation of the i^{th} component $\hat{x}_i$ of the estimator. Recall the formulation of $\mathbf{y}$ in (4.113). We will now formulate a similar expression for $\underline{\mathbf{y}}$. With the notation

$$\underline{\mathbf{h}}_i = \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_i^* \end{bmatrix} \in \mathbb{C}^{2N_\mathbf{y}}, \qquad \underline{\bar{\mathbf{H}}}_i = \begin{bmatrix} \bar{\mathbf{H}}_i \\ \bar{\mathbf{H}}_i^* \end{bmatrix} \in \mathbb{C}^{2N_\mathbf{y} \times (N_\mathbf{x}-1)} \tag{Z.1}$$

the augmented form of (4.113) yields

$$\underline{\mathbf{y}} = \underline{\mathbf{h}}_i x_i + \underline{\bar{\mathbf{H}}}_i \bar{\mathbf{x}}_i + \underline{\mathbf{n}}. \tag{Z.2}$$

Note that $\underline{\bar{\mathbf{H}}}_i$ is in fact not an augmented matrix. However, for the sake of simplicity and uniformity, we utilize the notation in (Z.1). Incorporating (Z.2) into (4.325) yields

$$\begin{aligned} E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] &= E_{\mathbf{y}|x_i}[\mathbf{e}_i^H \underline{\mathbf{y}} + b_i|x_i] && \text{(Z.3)} \\ &= E_{\mathbf{n},\bar{\mathbf{x}}_i|x_i}[\mathbf{e}_i^H(\underline{\mathbf{h}}_i x_i + \underline{\bar{\mathbf{H}}}_i \bar{\mathbf{x}}_i + \underline{\mathbf{n}}) + b_i|x_i] && \text{(Z.4)} \\ &= \mathbf{e}_i^H(\underline{\mathbf{h}}_i x_i + \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]) + b_i. && \text{(Z.5)} \end{aligned}$$

Because of the Gaussian assumption we have

$$E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H\Big(\underline{\mathbf{h}}_i x_i + \underline{\bar{\mathbf{H}}}_i\big(E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big(\sigma_{x_i}^2\big)^{-1}(x_i - E_{x_i}[x_i])\big)\Big) + b_i. \tag{Z.6}$$

By setting (Z.6) equal to x_i, one can see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if two conditions are fulfilled. The first condition is

$$\begin{aligned} \mathbf{e}_i^H \underline{\mathbf{h}}_i + \mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big(\sigma_{x_i}^2\big)^{-1} &= 1 && \text{(Z.7)} \\ \mathbf{e}_i^H\big(\underline{\mathbf{h}}_i \sigma_{x_i}^2 + \underline{\bar{\mathbf{H}}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big) &= \sigma_{x_i}^2. && \text{(Z.8)} \end{aligned}$$

The expression in the brackets in (Z.8) corresponds to $\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, which can be shown by

$$\begin{aligned} \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix} &= E_{\mathbf{y},\mathbf{x}}\Big[\big(\underline{\mathbf{y}} - E_\mathbf{y}[\underline{\mathbf{y}}]\big)\big(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]\big)^H\Big]\begin{bmatrix} 1 \\ 0 \end{bmatrix} && \text{(Z.9)} \\ &= E_{\mathbf{y},\mathbf{x}}\Big[\big(\underline{\mathbf{y}} - E_\mathbf{y}[\underline{\mathbf{y}}]\big)\Big(\begin{bmatrix} 1 & 0 \end{bmatrix}\big(\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i]\big)^H\Big)\Big] && \text{(Z.10)} \\ &= E_{\mathbf{y},\mathbf{x}}\Big[\big(\underline{\mathbf{h}}_i(x_i - E_{x_i}[x_i]) + \underline{\bar{\mathbf{H}}}_i(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \underline{\mathbf{n}}\big)(x_i - E_{x_i}[x_i])^H\Big] && \text{(Z.11)} \\ &= \underline{\mathbf{h}}_i \sigma_{x_i}^2 + \underline{\bar{\mathbf{H}}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}. && \text{(Z.12)} \end{aligned}$$

Hence, the first condition reads as

$$\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \sigma_{x_i}^2. \tag{Z.13}$$

The second condition follows from (Z.6) as

$$
\begin{aligned}
b_i =& -\mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i \Big(E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \mathbf{C}_{\bar{\mathbf{x}}_i x_i} \big(\sigma_{x_i}^2\big)^{-1} E_{x_i}[x_i] \Big) && \text{(Z.14)}\\
=& -\mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{e}_i^H \underbrace{\underline{\bar{\mathbf{H}}}_i \mathbf{C}_{\bar{\mathbf{x}}_i x_i}}_{\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix} - \underline{\mathbf{h}}_i \sigma_{x_i}^2} \big(\sigma_{x_i}^2\big)^{-1} E_{x_i}[x_i] && \text{(Z.15)}\\
=& -\mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \mathbf{e}_i^H \underline{\mathbf{h}}_i E_{x_i}[x_i] + \underbrace{\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix}1\\0\end{bmatrix}}_{\sigma_{x_i}^2} \big(\sigma_{x_i}^2\big)^{-1} E_{x_i}[x_i]) && \text{(Z.16)}\\
=& E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}], && \text{(Z.17)}
\end{aligned}
$$

where (Z.12) and (Z.13) were utilized. Eq. (Z.13) and (Z.17) allow to simplify the BMSE cost function $E_{\mathbf{y},\mathbf{x}}[|\hat{x}_i - x_i|^2]$ according to

$$
\begin{aligned}
J(\mathbf{e}_i) &= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \underline{\mathbf{y}} + b_i - x_i|^2] && \text{(Z.18)}\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - (x_i - E_{x_i}[x_i])|^2] && \text{(Z.19)}\\
&= E_{\mathbf{y},\mathbf{x}}[|\mathbf{e}_i^H \big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - \begin{bmatrix}1 & 0\end{bmatrix} (\underline{\mathbf{x}}_i - E_{x_i}[\underline{\mathbf{x}}_i])|^2] && \text{(Z.20)}\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i\mathbf{y}} \mathbf{e}_i + \begin{bmatrix}1 & 0\end{bmatrix} \underline{\mathbf{C}}_{x_i x_i} \begin{bmatrix}1\\0\end{bmatrix} && \text{(Z.21)}\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2 - \sigma_{x_i}^2 + \sigma_{x_i}^2 && \text{(Z.22)}\\
&= \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2. && \text{(Z.23)}
\end{aligned}
$$

Combining the cost function in (Z.23) and the constraint in (Z.13) leads to the optimization problem

$$
\mathbf{e}_{\mathrm{CWL},i} = \arg\min_{\mathbf{e}_i} \big(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2\big) \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix}1\\0\end{bmatrix} \big(\sigma_{x_i}^2\big)^{-1} = 1. \tag{Z.24}
$$

Note that this optimization problem equals the one for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}_{\mathbb{R}}$ in (4.336). Solving it will formally lead to the same expression for the CWCU WLMMSE estimator. However, a significant difference has been obtained. By making the assumption about an underlying linear model, the jointly Gaussian assumption of $\mathbf{x}$ and $\mathbf{y}_{\mathbb{R}}$ can be significantly relaxed. In fact, only the parameter vector $\mathbf{x}$ is required to be Gaussian for obtaining (4.336). The PDF of the noise $\mathbf{n}$ can be arbitrary. The only requirements on the noise vector are $E_{\mathbf{n}}[\mathbf{n}] = \mathbf{0}$, while $\mathbf{n}$ and $\mathbf{x}$ need to be uncorrelated.

For mutually independent parameters (case 3 of Result 4.3) it is possible to further relax the prerequisites on $\mathbf{x}$. In this case (Z.5) becomes

$$
E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = \mathbf{e}_i^H \underline{\mathbf{h}}_i x_i + \mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + b_i, \tag{Z.25}
$$

since $E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]$ is no longer dependent on x_i. By setting (Z.25) equal to x_i we see that the CWCU constraint $E_{\mathbf{y}|x_i}[\hat{x}_i|x_i] = x_i$ is fulfilled if

$$
\mathbf{e}_i^H \underline{\mathbf{h}}_i = 1 \tag{Z.26}
$$

and

$$
\begin{aligned}
b_i = &- \mathbf{e}_i^H \underline{\bar{\mathbf{H}}}_i E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && \text{(Z.27)}\\
= &- \mathbf{e}_i^H \big(E_{\mathbf{y}}[\underline{\mathbf{y}}] - \underline{\mathbf{h}}_i E_{x_i}[x_i] \big) && \text{(Z.28)}\\
= &\; E_{x_i}[x_i] - \mathbf{e}_i^H E_{\mathbf{y}}[\underline{\mathbf{y}}]. && \text{(Z.29)}
\end{aligned}
$$

Adapting (Z.12) for the case of mutually independent parameters shows that in this case

$$
\underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \underline{\mathbf{h}}_i \sigma_{x_i}^2 \tag{Z.30}
$$

holds. This result incorporated into (Z.26) yields

$$
\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \big(\sigma_{x_i}^2 \big)^{-1} = 1, \tag{Z.31}
$$

which equals (Z.13). Inserting in the BMSE cost function leads to the optimization problem

$$
\mathbf{e}_{\mathrm{CWL},i} = \arg\min_{\mathbf{e}_i} \; \big(\mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} \mathbf{e}_i - \sigma_{x_i}^2 \big) \quad \text{s.t.} \quad \mathbf{e}_i^H \underline{\mathbf{C}}_{\mathbf{y}x_i} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \big(\sigma_{x_i}^2 \big)^{-1} = 1. \tag{Z.32}
$$

This again equals the same optimization problem as for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}_{\mathbb{R}}$ in (4.336). As a consequence, the expressions for the CWCU WLMMSE estimator are formally the same.

AA Proof that (4.358) Holds for all Three Cases

We consider the real composite model for real-valued parameters

$$
\mathbf{y}_{\mathbb{R}} = \begin{bmatrix} \mathbf{y}_{\mathrm{R}} \\ \mathbf{y}_{\mathrm{I}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{\mathrm{R}} \\ \mathbf{H}_{\mathrm{I}} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_{\mathrm{R}} \\ \mathbf{n}_{\mathrm{I}} \end{bmatrix} = \mathbf{H}_{\mathbb{R}} \mathbf{x} + \mathbf{n}_{\mathbb{R}}. \tag{AA.1}
$$

Let $\mathbf{h}_{\mathbb{R},i}$ denote the i^{th} column of $\mathbf{H}_{\mathbb{R}}$, $\bar{\mathbf{H}}_{\mathbb{R},i}$ denote the matrix resulting from $\mathbf{H}_{\mathbb{R}}$ after deleting $\mathbf{h}_{\mathbb{R},i}$, x_i denote the i^{th} element of $\mathbf{x}$, and $\bar{\mathbf{x}}_i$ the vector resulting from $\mathbf{x}$ after deleting x_i. Then, the model in (AA.1) can be rewritten as

$$
\mathbf{y}_{\mathbb{R}} = \mathbf{h}_{\mathbb{R},i} x_i + \bar{\mathbf{H}}_{\mathbb{R},i} \bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}}. \tag{AA.2}
$$

Furthermore, it holds that

$$
\begin{aligned}
\mathbf{C}_{\mathbf{y}_{\mathbb{R}} x_i} &= E_{\mathbf{y}_{\mathbb{R}},\mathbf{x}}[(\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}}[\mathbf{y}_{\mathbb{R}}])(x_i - E_{x_i}[x_i])^*] && \text{(AA.3)}\\
&= E_{\mathbf{y}_{\mathbb{R}},\mathbf{x}}\big[\big(\mathbf{h}_{\mathbb{R},i}(x_i - E_{x_i}[x_i]) + \bar{\mathbf{H}}_{\mathbb{R},i}(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}_{\mathbb{R}} \big)(x_i - E_{x_i}[x_i])^* \big] && \text{(AA.4)}\\
&= \mathbf{h}_{\mathbb{R},i} \sigma_{x_i}^2 + \bar{\mathbf{H}}_{\mathbb{R},i} \mathbf{C}_{\bar{\mathbf{x}}_i x_i}. && \text{(AA.5)}
\end{aligned}
$$

For the case of Gaussian and real-valued $\mathbf{x}$, it holds that

$$\begin{aligned}
E_{\mathbf{y}_\mathbb{R}|x_i}[\mathbf{y}_\mathbb{R}|x_i] &= E_{\mathbf{y}|x_i}\left[\mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{n}_\mathbb{R}|x_i\right] && \text{(AA.6)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i|x_i] && \text{(AA.7)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}\big(E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma^2_{x_i})^{-1}(x_i - E_{x_i}[x_i])\big) && \text{(AA.8)}\\
&= \big(\mathbf{h}_{\mathbb{R},i} + \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma^2_{x_i})^{-1}\big)x_i + \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma^2_{x_i})^{-1}E_{x_i}[x_i] && \text{(AA.9)}\\
&= \underbrace{\big(\mathbf{h}_{\mathbb{R},i}\sigma^2_{x_i} + \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big)}_{\mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}}(\sigma^2_{x_i})^{-1}x_i + E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] - \mathbf{h}_{\mathbb{R},i}E_{x_i}[x_i]\\
&\quad - \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma^2_{x_i})^{-1}E_{x_i}[x_i] && \text{(AA.10)}\\
&= \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}(\sigma^2_{x_i})^{-1}x_i + E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] - \underbrace{\big(\mathbf{h}_{\mathbb{R},i}\sigma^2_{x_i} + \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big)}_{\mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}}(\sigma^2_{x_i})^{-1}E_{x_i}[x_i] && \text{(AA.11)}\\
&= \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}(\sigma^2_{x_i})^{-1}x_i + E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] - \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}(\sigma^2_{x_i})^{-1}E_{x_i}[x_i] && \text{(AA.12)}\\
&= E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] + \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}(\sigma^2_{x_i})^{-1}(x_i - E_{x_i}[x_i]). && \text{(AA.13)}
\end{aligned}$$

For the third case we assumed mutually independent but otherwise arbitrary distributed real-valued elements of $\mathbf{x}$. Then it holds that

$$\begin{aligned}
E_{\mathbf{y}_\mathbb{R}|x_i}[\mathbf{y}_\mathbb{R}|x_i] &= E_{\mathbf{y}_\mathbb{R}|x_i}\left[\mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] + \mathbf{n}_\mathbb{R}|x_i\right] && \text{(AA.14)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && \text{(AA.15)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] - \mathbf{h}_{\mathbb{R},i}E_{x_i}[x_i] && \text{(AA.16)}\\
&= E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] + \underbrace{\mathbf{h}_{\mathbb{R},i}\sigma^2_{x_i}}_{\mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}}(\sigma^2_{x_i})^{-1}(x_i - E_{x_i}[x_i]) && \text{(AA.17)}\\
&= E_{\mathbf{y}_\mathbb{R}}[\mathbf{y}_\mathbb{R}] + \mathbf{C}_{\mathbf{y}_\mathbb{R}x_i}(\sigma^2_{x_i})^{-1}(x_i - E_{x_i}[x_i]), && \text{(AA.18)}
\end{aligned}$$

concluding the proof. □

AB Proof that (4.360) Holds for all Three Cases

The proof of (4.360) for jointly Gaussian $\mathbf{x}$ and $\mathbf{y}_\mathbb{R}$ is a straightforward extension of the consideration in Appendix N.

For the second case with real-valued and Gaussian $\mathbf{x}$, we first rewrite $\mathbf{y}_\mathbb{R} - E_{\mathbf{y}_\mathbb{R}|x_i}[\mathbf{y}_\mathbb{R}|x_i]$

utilizing (AA.2) and (4.358) as

$$\begin{aligned}
\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i] &= \mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}}[\mathbf{y}_{\mathbb{R}}] - \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.1)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \mathbf{h}_{\mathbb{R},i}E_{x_i}[x_i] - \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.2)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \mathbf{h}_{\mathbb{R},i}E_{x_i}[x_i] - \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) - \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.3)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.4)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \bar{\mathbf{H}}_{\mathbb{R},i}\underbrace{\left(E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] - \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i])\right)}_{E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]} && \text{(AB.5)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}\big(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]\big) + \mathbf{n}_{\mathbb{R}}. && \text{(AB.6)}
\end{aligned}$$

With that, the conditional covariance matrix can be derived as

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}|x_i} &= E_{\mathbf{y}_{\mathbb{R}}|x_i}\Big[\big(\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i]\big)\big(\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i]\big)^H|x_i\Big] && \text{(AB.7)}\\
&= E_{\mathbf{y}_{\mathbb{R}}|x_i}\Big[\big(\bar{\mathbf{H}}_{\mathbb{R},i}\big(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]\big) + \mathbf{n}_{\mathbb{R}}\big)\big(\bar{\mathbf{H}}_{\mathbb{R},i}\big(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i|x_i}[\bar{\mathbf{x}}_i|x_i]\big) + \mathbf{n}_{\mathbb{R}}\big)^H|x_i\Big] && \text{(AB.8)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i|x_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H + \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}}. && \text{(AB.9)}
\end{aligned}$$

Incorporating (AA.5) and the fact that $\mathbf{x}$ is real-valued and Gaussian yields

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}|x_i} &= \bar{\mathbf{H}}_{\mathbb{R},i}\big(\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i} - \mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\big)\bar{\mathbf{H}}_{\mathbb{R},i}^H + \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}} && \text{(AB.10)}\\
&= \underbrace{\bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H + \mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2\mathbf{h}_{\mathbb{R},i}^H + \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\mathbf{h}_{\mathbb{R},i}^H + \mathbf{h}_{\mathbb{R},i}\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H + \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}}}_{\mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}}}\\
&\quad - \mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2\mathbf{h}_{\mathbb{R},i}^H - \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\mathbf{h}_{\mathbb{R},i}^H\\
&\quad - \mathbf{h}_{\mathbb{R},i}\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H - \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H && \text{(AB.11)}\\
&= \mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}} - \underbrace{\big(\mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2 + \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i x_i}\big)}_{\mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}}(\sigma_{x_i}^2)^{-1}\underbrace{\big(\sigma_{x_i}^2\mathbf{h}_{\mathbb{R},i}^H + \mathbf{C}_{x_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H\big)}_{\mathbf{C}_{x_i\mathbf{y}_{\mathbb{R}}}} && \text{(AB.12)}\\
&= \mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}} - \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}_{\mathbb{R}}}. && \text{(AB.13)}
\end{aligned}$$

For the third case where the elements of $\mathbf{x}$ are uncorrelated, $\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i]$ follows as

$$\begin{aligned}
\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i] &= \mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}}[\mathbf{y}_{\mathbb{R}}] - \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.14)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \mathbf{h}_{\mathbb{R},i}E_{x_i}[x_i] - \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.15)}\\
&= \mathbf{h}_{\mathbb{R},i}x_i + \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \mathbf{h}_{\mathbb{R},i}E_{x_i}[x_i] - \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]\\
&\quad - \mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2(\sigma_{x_i}^2)^{-1}(x_i - E_{x_i}[x_i]) && \text{(AB.16)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}\bar{\mathbf{x}}_i + \mathbf{n}_{\mathbb{R}} - \bar{\mathbf{H}}_{\mathbb{R},i}E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i] && \text{(AB.17)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}_{\mathbb{R}}. && \text{(AB.18)}
\end{aligned}$$

Incorporating (AA.5) allows deriving the conditional covariance matrix as

$$\begin{aligned}
\mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}|x_i} &= E_{\mathbf{y}_{\mathbb{R}}|x_i}\left[\left(\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i]\right)\left(\mathbf{y}_{\mathbb{R}} - E_{\mathbf{y}_{\mathbb{R}}|x_i}[\mathbf{y}_{\mathbb{R}}|x_i]\right)^H|x_i\right] && \text{(AB.19)}\\
&= E_{\mathbf{y}_{\mathbb{R}}|x_i}\left[\left(\bar{\mathbf{H}}_{\mathbb{R},i}(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}_{\mathbb{R}}\right)\left(\bar{\mathbf{H}}_{\mathbb{R},i}(\bar{\mathbf{x}}_i - E_{\bar{\mathbf{x}}_i}[\bar{\mathbf{x}}_i]) + \mathbf{n}_{\mathbb{R}}\right)^H|x_i\right] && \text{(AB.20)}\\
&= \bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H + \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}} && \text{(AB.21)}\\
&= \underbrace{\bar{\mathbf{H}}_{\mathbb{R},i}\mathbf{C}_{\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i}\bar{\mathbf{H}}_{\mathbb{R},i}^H + \mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2\mathbf{h}_{\mathbb{R},i}^H + \mathbf{C}_{\mathbf{n}_{\mathbb{R}}\mathbf{n}_{\mathbb{R}}}}_{\mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}}} - \mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2\mathbf{h}_{\mathbb{R},i}^H && \text{(AB.22)}\\
&= \mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}} - \underbrace{\mathbf{h}_{\mathbb{R},i}\sigma_{x_i}^2}_{\mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}}(\sigma_{x_i}^2)^{-1}\underbrace{\sigma_{x_i}^2\mathbf{h}_{\mathbb{R},i}^H}_{\mathbf{C}_{x_i\mathbf{y}_{\mathbb{R}}}} && \text{(AB.23)}\\
&= \mathbf{C}_{\mathbf{y}_{\mathbb{R}}\mathbf{y}_{\mathbb{R}}} - \mathbf{C}_{\mathbf{y}_{\mathbb{R}}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\mathbf{y}_{\mathbb{R}}}. && \text{(AB.24)}
\end{aligned}$$

□

AC Derivation of the Conditional Properties of the WLMMSE Estimator for Real-Valued Parameters

Consider the WLMMSE for x_i in (4.49). For this estimator, the conditional mean follows as

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] = E_{x_i}[x_i] + \begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\left(E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] - E_{\mathbf{y}}[\underline{\mathbf{y}}]\right). \tag{AC.1}$$

Inserting (4.359) produces

$$\begin{aligned}
E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] &= E_{x_i}[x_i] + \underbrace{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma_{x_i}^2}}_{[\mathbf{D}]_{i,i}^{-1}}(x_i - E_{x_i}[x_i]) && \text{(AC.2)}\\
&= E_{x_i}[x_i] + [\mathbf{D}]_{i,i}^{-1}(x_i - E_{x_i}[x_i]) && \text{(AC.3)}\\
&= [\mathbf{D}]_{i,i}^{-1}x_i + (1 - [\mathbf{D}]_{i,i}^{-1})E_{x_i}[x_i]. && \text{(AC.4)}
\end{aligned}$$

Now, the conditional bias can be derived as

$$b(\hat{x}_{\mathrm{WL},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i] - x_i \tag{AC.5}$$
$$= \Big([\mathbf{D}]^{-1}_{i,i} - 1\Big)x_i - \Big([\mathbf{D}]^{-1}_{i,i} - 1\Big)E_{x_i}[x_i] \tag{AC.6}$$
$$= \Big([\mathbf{D}]^{-1}_{i,i} - 1\Big)(x_i - E_{x_i}[x_i]). \tag{AC.7}$$

The conditional variance follows with (AC.1) as

$$\mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) = E_{\mathbf{y}|x_i}\Big[\big|\hat{x}_{\mathrm{WL},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{WL},i}|x_i]\big|^2|x_i\Big] \tag{AC.8}$$
$$= E_{\mathbf{y}|x_i}\Big[\Big|E_{x_i}[x_i] + \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big) - E_{x_i}[x_i]$$
$$- \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\big(E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\Big|^2|x_i\Big] \tag{AC.9}$$
$$= E_{\mathbf{y}|x_i}\left[\left|\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\big(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\big)\right|^2|x_i\right] \tag{AC.10}$$
$$= \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\underline{\mathbf{C}}_{\underline{\mathbf{y}}\underline{\mathbf{y}}|x_i}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i}. \tag{AC.11}$$

Utilizing (4.361) allows

$$\mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) = \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\left(\underline{\mathbf{C}}_{\underline{\mathbf{y}}\underline{\mathbf{y}}} - \mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma^2_{x_i}}\mathbf{C}_{x_i\underline{\mathbf{y}}}\right)\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i} \tag{AC.12}$$
$$= \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i} - \mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma^2_{x_i}}\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i} \tag{AC.13}$$
$$= \underbrace{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma^2_{x_i}}}_{[\mathbf{D}]^{-1}_{i,i}}\sigma^2_{x_i} - \underbrace{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma^2_{x_i}}}_{[\mathbf{D}]^{-1}_{i,i}}\underbrace{\mathbf{C}_{x_i\underline{\mathbf{y}}}\underline{\mathbf{C}}^{-1}_{\underline{\mathbf{y}}\underline{\mathbf{y}}}\mathbf{C}_{\underline{\mathbf{y}}x_i}\frac{1}{\sigma^2_{x_i}}}_{[\mathbf{D}]^{-1}_{i,i}}\sigma^2_{x_i} \tag{AC.14}$$
$$= [\mathbf{D}]^{-1}_{i,i}\sigma^2_{x_i} - [\mathbf{D}]^{-1}_{i,i}[\mathbf{D}]^{-1}_{i,i}\sigma^2_{x_i} \tag{AC.15}$$
$$= [\mathbf{D}]^{-1}_{i,i}\Big(1 - [\mathbf{D}]^{-1}_{i,i}\Big)\sigma^2_{x_i}. \tag{AC.16}$$

With the conditional variance, the conditional bias, and the fact that $[\mathbf{D}]_{i,i} \geq 1$ is real-valued, the conditional MSE can be derived as

$$\mathrm{mse}(\hat{x}_{\mathrm{WL},i}|x_i) = E_{\mathbf{y}|x_i}\Big[|\hat{x}_{\mathrm{WL},i} - x_i|^2|x_i\Big] \tag{AC.17}$$
$$= \mathrm{var}(\hat{x}_{\mathrm{WL},i}|x_i) + |b(\hat{x}_{\mathrm{WL},i}|x_i)|^2 \tag{AC.18}$$
$$= [\mathbf{D}]^{-1}_{i,i}\Big(1 - [\mathbf{D}]^{-1}_{i,i}\Big)\sigma^2_{x_i} + \Big|\Big([\mathbf{D}]^{-1}_{i,i} - 1\Big)(x_i - E_{x_i}[x_i])\Big|^2 \tag{AC.19}$$
$$= [\mathbf{D}]^{-1}_{i,i}\Big(1 - [\mathbf{D}]^{-1}_{i,i}\Big)\sigma^2_{x_i} + \Big(1 - [\mathbf{D}]^{-1}_{i,i}\Big)^2|(x_i - E_{x_i}[x_i])|^2. \tag{AC.20}$$

AD Derivation of the Conditional Properties of the CWCU WLMMSE Estimator for Real-Valued Parameters

The conditional mean and the conditional bias of the CWCU WLMMSE given in (4.92) has the conditional mean

$$E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] = x_i \tag{AD.1}$$

and the conditional bias

$$b(\hat{x}_{\mathrm{CWL},i}|x_i) = E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i] - x_i = 0 \tag{AD.2}$$

since it fulfills the CWCU constraints. The conditional variance follows with (4.343) as

$$\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) &= E_{\mathbf{y}|x_i}\left[\left|\hat{x}_{\mathrm{CWL},i} - E_{\mathbf{y}|x_i}[\hat{x}_{\mathrm{CWL},i}|x_i]\right|^2 |x_i\right] && \text{(AD.3)}\\
&= E_{\mathbf{y}|x_i}\Bigg[\Bigg|E_{x_i}[x_i] + \mathbf{e}_{\mathrm{CWL},i}^H\big(\underline{\mathbf{y}} - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\\
&\qquad - E_{x_i}[x_i] - \mathbf{e}_{\mathrm{CWL},i}^H\big(E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i] - E_{\mathbf{y}}[\underline{\mathbf{y}}]\big)\Bigg|^2 |x_i\Bigg] && \text{(AD.4)}\\
&= E_{\mathbf{y}|x_i}\left[\left|\mathbf{e}_{\mathrm{CWL},i}^H\underline{\mathbf{y}} - \mathbf{e}_{\mathrm{CWL},i}^H E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\right|^2 |x_i\right] && \text{(AD.5)}\\
&= E_{\mathbf{y}|x_i}\left[\left|\mathbf{e}_{\mathrm{CWL},i}^H\big(\underline{\mathbf{y}} - E_{\mathbf{y}|x_i}[\underline{\mathbf{y}}|x_i]\big)\right|^2 |x_i\right] && \text{(AD.6)}\\
&= \mathbf{e}_{\mathrm{CWL},i}^H \underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}|x_i}\mathbf{e}_{\mathrm{CWL},i}. && \text{(AD.7)}
\end{aligned}$$

Utilizing (4.361), (4.341) and (4.347), (AD.7) reads as

$$\begin{aligned}
\mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) &= \mathbf{e}_{\mathrm{CWL},i}^H\Big(\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\underline{\mathbf{y}}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\underline{\mathbf{y}}}\Big)\mathbf{e}_{\mathrm{CWL},i} && \text{(AD.8)}\\
&= \left(\frac{\sigma_{x_i}^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}}\right)^2\\
&\quad\times\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\Big(\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}} - \mathbf{C}_{\underline{\mathbf{y}}x_i}(\sigma_{x_i}^2)^{-1}\mathbf{C}_{x_i\underline{\mathbf{y}}}\Big)\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{C}_{\underline{\mathbf{y}}x_i}\begin{bmatrix}1\\0\end{bmatrix} && \text{(AD.9)}\\
&= \frac{\left(\sigma_{x_i}^2\right)^2}{\begin{bmatrix}1 & 0\end{bmatrix}\underline{\mathbf{C}}_{x_i\mathbf{y}}\underline{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1}\underline{\mathbf{C}}_{\mathbf{y}x_i}\begin{bmatrix}1\\0\end{bmatrix}} - \sigma_{x_i}^2 && \text{(AD.10)}\\
&= \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). && \text{(AD.11)}
\end{aligned}$$

Finally, the conditional MSE can be derived as

$$\begin{aligned}
\mathrm{mse}(\hat{x}_{\mathrm{CWL},i}|x_i) &= E_{\mathbf{y}|x_i}\left[|\hat{x}_{\mathrm{CWL},i} - x_i|^2|x_i\right] && \text{(AD.12)}\\
&= \mathrm{var}(\hat{x}_{\mathrm{CWL},i}|x_i) && \text{(AD.13)}\\
&= \sigma_{x_i}^2([\mathbf{D}]_{i,i} - 1). && \text{(AD.14)}
\end{aligned}$$

AE Complexity Analysis of the LMS Algorithm for Real-Valued Filter Coefficients

We analyze the computational complexity in terms of the required multiplications and additions for the update step in Result 5.1.

- Derive $e_k \in \mathbb{C}$
  $e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$	$2N_\mathbf{w}$	$2N_\mathbf{w} - 2$
$y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$	0	2
e_k summary	$2N_\mathbf{w}$	$2N_\mathbf{w}$

- Derive $\mathbf{w}_k \in \mathbb{R}^{N_\mathbf{w}}$
  $\mathbf{w}_k = \mathbf{w}_{k-1} + \mu \mathrm{Re}\{\mathbf{x}_k^* e_k\}$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\mathrm{Re}\{\mathbf{x}_k^* e_k\} \in \mathbb{R}^{N_\mathbf{w}}$	$2N_\mathbf{w}$	$N_\mathbf{w}$
$\mu \mathrm{Re}\{\mathbf{x}_k^* e_k\} \in \mathbb{R}^{N_\mathbf{w}}$	$N_\mathbf{w}$	0
$\mathbf{w}_{k-1} + \mu \mathrm{Re}\{\mathbf{x}_k^* e_k\} \in \mathbb{R}^{N_\mathbf{w}}$	0	$N_\mathbf{w}$
$\mathbf{w}_k$ summary	$3N_\mathbf{w}$	$2N_\mathbf{w}$

- Total

Operation	Amount
Real-Valued Multiplications	$5N_\mathbf{w}$
Real-Valued Additions	$4N_\mathbf{w}$

AF Complexity Analysis of the RLS Algorithm for Real-Valued Filter Coefficients

Here, we analyze the computational complexity in terms of the required multiplications and additions for the update step in Result 5.2.

- Derive $\mathbf{G}_k \in \mathbb{C}^{N_\mathbf{w}\times 2}$

  $\mathbf{G}_k = \lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big)^{-1}$

Expression	Real-Valued Multiplications	Real-Valued Additions	Real-Valued Divisions
$\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k \in \mathbb{C}^{N_\mathbf{w}\times 2}$	$4N_\mathbf{w}^2$	$4N_\mathbf{w}^2 - 4N_\mathbf{w}$	0
$\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k \in \mathbb{C}^{2\times 2}$ (Hermitian matrix)	$16N_\mathbf{w}$	$16N_\mathbf{w} - 8$	0
$\lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k \in \mathbb{C}^{2\times 2}$	6	0	0
$\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k \in \mathbb{C}^{2\times 2}$	0	2	0
$\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big)^{-1} \in \mathbb{C}^{2\times 2}$	3	1	1
$\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big)^{-1} \in \mathbb{C}^{N_\mathbf{w}\times 2}$	$16N_\mathbf{w}$	$12N_\mathbf{w}$	0
$\lambda^{-1}\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big(\mathbf{I}^{2\times 2} + \lambda^{-1}\mathbf{X}_k^H\widetilde{\mathbf{P}}_{k-1}\mathbf{X}_k\Big)^{-1} \in \mathbb{C}^{N_\mathbf{w}\times 2}$	$4N_\mathbf{w}$	0	0
$\mathbf{G}_k$ summary	$4N_\mathbf{w}^2 + 36N_\mathbf{w} + 9$	$4N_\mathbf{w}^2 + 24N_\mathbf{w} - 4$	1

- Derive $\underline{\mathbf{e}}_k \in \mathbb{C}^2$
  Is the augmented version of $e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$	$2N_\mathbf{w}$	$2(N_\mathbf{w} - 1)$
$y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$	0	2
$\underline{\mathbf{e}}_k$ summary	$2N_\mathbf{w}$	$2N_\mathbf{w}$

- Derive $\widetilde{\mathbf{P}}_k \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$
  $\widetilde{\mathbf{P}}_k = \lambda^{-1}\left(\widetilde{\mathbf{P}}_{k-1} - \mathbf{G}_k \mathbf{X}_k^H \widetilde{\mathbf{P}}_{k-1}\right)$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\mathbf{X}_k^H \widetilde{\mathbf{P}}_{k-1} \in \mathbb{C}^{2 \times N_\mathbf{w}}$ results from the update step for $\mathbf{G}_k$	0	0
$\mathbf{G}_k \mathbf{X}_k^H \widetilde{\mathbf{P}}_{k-1} \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$ (symmetric matrix)	$4N_\mathbf{w}^2$	$4N_\mathbf{w}^2$
$\widetilde{\mathbf{P}}_{k-1} - \mathbf{G}_k \mathbf{X}_k^H \widetilde{\mathbf{P}}_{k-1} \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$	0	$N_\mathbf{w}^2$
$\lambda^{-1}\left(\widetilde{\mathbf{P}}_{k-1} - \mathbf{G}_k \mathbf{X}_k^H \widetilde{\mathbf{P}}_{k-1}\right) \in \mathbb{R}^{N_\mathbf{w} \times N_\mathbf{w}}$	$N_\mathbf{w}$	0
$\widetilde{\mathbf{P}}_k$ summary	$5N_\mathbf{w}^2$	$5N_\mathbf{w}^2$

- Derive $\mathbf{w}_k \in \mathbb{C}^{N_\mathbf{w}}$
  $\mathbf{w}_k^T = \mathbf{w}_{k-1}^T + \mathbf{e}_k^T \mathbf{G}_k^H$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\mathbf{e}_k^T \mathbf{G}_k^H \in \mathbb{R}^{N_\mathbf{w}}$	$2N_\mathbf{w}$	$2N_\mathbf{w}$
$\mathbf{w}_{k-1}^T + \mathbf{e}_k^T \mathbf{G}_k^H \in \mathbb{R}^{N_\mathbf{w}}$	0	$N_\mathbf{w}$
$\mathbf{w}_k$ summary	$2N_\mathbf{w}$	$3N_\mathbf{w}$

- Total

Operation	Amount
Real-Valued Multiplications	$9N_{\mathbf{w}}^2 + 40N_{\mathbf{w}} + 9$
Real-Valued Additions	$9N_{\mathbf{w}}^2 + 29N_{\mathbf{w}} - 4$
Real-Valued Divisions	1

AG Complexity Analysis of the Bayesian NLMS Algorithm

An analysis of the computational complexity in terms of the required multiplications, divisions and additions for the update step in Result 5.3 is presented in this appendix.

- Derive $e_k \in \mathbb{C}$
  $e_k = y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$	$4N_{\mathbf{w}}$	$2N_{\mathbf{w}} - 2$
$y_k - \mathbf{w}_{k-1}^T \mathbf{x}_k \in \mathbb{C}$	0	2
e_k summary	$4N_{\mathbf{w}}$	$2N_{\mathbf{w}}$

- Derive normalizing term $\frac{1}{\epsilon + \frac{1}{\sigma_n^2}\|\mathbf{x}_k\|_2^2 + a_k \lambda_1\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)} \in \mathbb{R}$

Expression	Real-Valued Multiplications	Real-Valued Additions	Real-Valued Divisions
$\|\mathbf{x}_k\|_2^2 \in \mathbb{R}$	$2N_{\mathbf{w}}$	$2N_{\mathbf{w}} - 1$	0
$\frac{1}{\epsilon + \frac{1}{\sigma_n^2}\|\mathbf{x}_k\|_2^2 + a_k \lambda_1\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)} \in \mathbb{R}$	2	2	1
$\mathbf{w}_k$ summary	$2N_{\mathbf{w}} + 2$	$2N_{\mathbf{w}} + 1$	1

- Derive $\mathbf{w}_k \in \mathbb{C}^{N_{\mathbf{w}}}$
  $\mathbf{w}_k = \mathbf{w}_{k-1} + \mu_{n,k} \frac{1}{\epsilon + \frac{1}{\sigma_n^2}\|\mathbf{x}_k\|_2^2 + a_k \lambda_1\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)} \left(\frac{1}{\sigma_n^2} \mathbf{x}_k^* e_k - a_k \mathbf{C}_{\mathbf{hh}}^{-1} (\mathbf{w}_{k-1} - E_{\mathbf{h}}[\mathbf{h}]) \right)$

Expression	Real-Valued Multiplications	Real-Valued Additions
$\frac{1}{\sigma_n^2}\mathbf{x}_k^* e_k \in \mathbb{C}^{N_\mathbf{w}}$	$6N_\mathbf{w}$	$2N_\mathbf{w}$
$\mathbf{w}_{k-1} - E_\mathbf{h}[\mathbf{h}] \in \mathbb{C}^{N_\mathbf{w}}$	0	$2N_\mathbf{w}$
$a_k \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_\mathbf{h}[\mathbf{h}]) \in \mathbb{C}^{N_\mathbf{w}}$	$6N_\mathbf{w}$	$2N_\mathbf{w}$
$\frac{1}{\sigma_n^2}\mathbf{x}_k^* e_k - a_k \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_\mathbf{h}[\mathbf{h}]) \in \mathbb{C}^{N_\mathbf{w}}$	0	$2N_\mathbf{w}$
$\mathbf{w}_{k-1} + \mu_{n,k} \frac{1}{\epsilon + \frac{1}{\sigma_n^2}\Vert\mathbf{x}_k\Vert_2^2 + a_k \lambda_1\left(\mathbf{C}_{\mathbf{hh}}^{-1}\right)}$ $\times\left(\frac{1}{\sigma_n^2}\mathbf{x}_k^* e_k - a_k \mathbf{C}_{\mathbf{hh}}^{-1}(\mathbf{w}_{k-1} - E_\mathbf{h}[\mathbf{h}])\right) \in \mathbb{C}^{N_\mathbf{w}}$	$4N_\mathbf{w}$	$2N_\mathbf{w}$
$\mathbf{w}_k$ summary	$16N_\mathbf{w}$	$10N_\mathbf{w}$

- Total

Operation	Amount
Real-Valued Multiplications	$22N_\mathbf{w} + 2$
Real-Valued Additions	$14N_\mathbf{w} + 1$
Real-Valued Divisions	1

List of Examples

List of Abbreviations

ALS approximate least squares

AWGN additive white Gaussian noise

BER bit error ratio

BLUE best linear unbiased estimator

BMSE Bayesian mean square error

BWLUE best widely linear unbiased estimator

CIR channel impulse response

CWCU component-wise conditionally unbiased

CWCU LMMSE component-wise conditionally unbiased linear minimum mean square error

CWCU WLMMSE component-wise conditionally unbiased widely linear minimum mean square error

DC direct current

DFT discrete Fourier transform

EM expectation-maximization

FFT fast Fourier transform

FIR finite impulse response

IDFT inverse discrete Fourier transform

IFFT inverse fast Fourier transform

i.i.d. independent and identically distributed

LLR log-likelihood ratio

LMMSE linear minimum mean square error

LMS least mean square

LS least squares

LTI linear time-invariant

MAP maximum a posteriori

MIMO multiple input multiple output

ML maximum likelihood

ML-EM maximum likelihood expectation-maximization

MMSE minimum mean square error

MSE mean square error

MVDR minimum variance distortionless response

MVU minimum variance unbiased

NLMS normalized least mean squares

OFDM orthogonal frequency division multiplexing

PDF probability density function

PWCU part-wise conditionally unbiased

PWCU WLMMSE part-wise conditionally unbiased widely linear minimum mean square error

QAM quadrature amplitude modulation

QPSK quadrature phase-shift keying

RLS recursive least squares

STLN structured total least norm

STLS structured total least squares

TLS total least squares

UW-OFDM unique-word orthogonal frequency division multiplexing

WLAN wireless local area network

WLLS widely linear least squares

WLMMSE widely linear minimum mean square error

WLS weighted least squares

WWLLS weighted widely linear least squares

Bibliography

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1993, Vol. 1.

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

[3] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," In *IEEE Transactions on Signal Processing*, Vol. 59, No. 11, pp. 5101–5125, Nov 2011.

[4] W. Wirtinger, "Zur formalen Theorie der Funktionen von mehr komplexen Veränderlichen," In *Mathematische Annalen*, Vol. 97, No. 1, pp. 357–375, 1927.

[5] C. W. Therrien, *Discrete random signals and statistical signal processing.* Prentice Hall PTR, 1992.

[6] K. Kreutz-Delgado, "The Complex Gradient Operator and the CR-Calculus," In *ArXiv e-prints arXiv:0906.4835 [math.OC]*, June 2009.

[7] M. Pilanci, O. Arikan, and M. C. Pinar, "Structured Least Squares Problems and Robust Estimators," In *IEEE Transactions on Signal Processing*, Vol. 58, No. 5, pp. 2453–2465, May 2010.

[8] S. V. Huffel, H. Park, and J. B. Rosen, "Formulation and solution of structured total least norm problems for parameter estimation," In *IEEE Transactions on Signal Processing*, Vol. 44, No. 10, pp. 2464–2474, Oct 1996.

[9] A. Kukush, I. Markovsky, and S. V. Huffel, "Consistency of the structured total least squares estimator in a multivariate errors-in-variables model," In *Journal of Statistical Planning and Inference*, Vol. 133, No. 2, pp. 315 – 358, 2005.

[10] B. D. Moor, "Structured total least squares and L2 approximation problems," In *Linear Algebra and its Applications*, Vol. 188, pp. 163 – 205, 1993.

[11] G. H. Golub and C. F. van Loan, "An Analysis of the Total Least Squares Problem," In *SIAM Journal on Numerical Analysis*, Vol. 17, No. 6, pp. 883–893, 1980.

[12] I. Markovsky and S. V. Huffel, "Overview of total least-squares methods," In *Signal Processing*, Vol. 87, No. 10, pp. 2283 – 2302, 2007.

[13] M. Triki and D. T. M. Slock, "Component-Wise Conditionally Unbiased Bayesian Parameter Estimation: General Concept and Applications to Kalman Filtering and

LMMSE Channel Estimation," In *Proc. of the 39th Asilomar Conference on Signals, Systems and Computers*, Oct 2005, pp. 670–674.

[14] M. Triki, A. Salah, and D. T. M. Slock, "Interference cancellation with Bayesian channel models and application to TDOA/IPDL mobile positioning," In *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, 2005.*, Vol. 1, Aug 2005, pp. 299–302.

[15] M. Triki and D. T. M. Slock, "Investigation of Some Bias and MSE Issues in Block-Component-Wise Conditionally Unbiased LMMSE," In *Prof. of the Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 1420–1424.

[16] A. Elmaghraby, *Transmitter Leakage Cancellation in Cellular Handset Receivers.* Dr. Hut, 2016.

[17] E. Ollila, "On the Circularity of a Complex Random Variable," In *IEEE Signal Processing Letters*, Vol. 15, pp. 841–844, 2008.

[18] M. Novey, T. Adali, and A. Roy, "Circularity and Gaussianity Detection Using the Complex Generalized Gaussian Distribution," In *IEEE Signal Processing Letters*, Vol. 16, No. 11, pp. 993–996, Nov 2009.

[19] J. Eriksson, E. Ollila, and V. Koivunen, "Essential Statistics and Tools for Complex Random Variables," In *IEEE Transactions on Signal Processing*, Vol. 58, No. 10, pp. 5400–5408, Oct 2010.

[20] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," In *IEEE Transactions on Signal Processing*, Vol. 51, No. 3, pp. 714–725, Mar 2003.

[21] B. Picinbono, "On circularity," In *IEEE Transactions on Signal Processing*, Vol. 42, No. 12, pp. 3473–3482, Dec 1994.

[22] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," In *IEEE transactions on Signal Processing*, Vol. 43, No. 8, pp. 2030–2033, 1995.

[23] M. Novey, T. Adali, and A. Roy, "A Complex Generalized Gaussian Distribution: Characterization, Generation, and Estimation," In *IEEE Transactions on Signal Processing*, Vol. 58, No. 3, pp. 1427–1433, Mar 2010.

[24] B. Picinbono, "Second-order complex random vectors and normal distributions," In *IEEE Transactions on Signal Processing*, Vol. 44, No. 10, pp. 2637–2640, 1996.

[25] A. van den Bos, "The multivariate complex normal distribution-a generalization," In *IEEE Transactions on Information Theory*, Vol. 41, No. 2, pp. 537–539, Mar 1995.

[26] A. van den Bos, "Complex gradient and Hessian," In *IEE Proceedings - Vision, Image and Signal Processing*, Vol. 141, No. 6, pp. 380–383, Dec 1994.

[27] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," In *IEE Proceedings F: Communications Radar and Signal Processing*, Vol. 130, 1983, pp. 11–16.

[28] A. Hjorungnes and D. Gesbert, "Complex-Valued Matrix Differentiation: Techniques and Key Results," In *IEEE Transactions on Signal Processing*, Vol. 55, No. 6, pp. 2740–2746, June 2007.

[29] D. P. Bertsekas, *Nonlinear programming.* Athena Scientific, 1999.

[30] E. K. Chong and S. H. Zak, *An introduction to optimization.* John Wiley & Sons, 2013, Vol. 4.

[31] O. Lang, V. Elvira, and M. Huemer, "Estimation of Real Valued Impulse Responses based on Noisy Magnitude and Phase Measurements," In *Proc. of the 51th Asilomar Conference on Signals, Systems and Computers (to be published)*, Nov 2017.

[32] T. Adali and S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*, ser. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2010.

[33] P. Chevalier, J.-P. Delmas, and A. Oukaci, "Properties, performance and practical interest of the widely linear MMSE beamformer for nonrectilinear signals," In *Signal Processing*, Vol. 97, pp. 269 – 281, 2014.

[34] P. Chevalier, "Optimal time invariant and widely linear spatial filtering for radiocommunications," In *Proc. of the 8th European Signal Processing Conference (EUSIPCO 1996)*, Sept 1996, pp. 1–4.

[35] H. Gerstacker, R. Schober, and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," In *IEEE Transactions on Communications*, Vol. 51, No. 9, pp. 1512–1523, Sept 2003.

[36] P. Chevalier and F. Pipon, "New insights into optimal widely linear array receivers for the demodulation of BPSK, MSK, and GMSK signals corrupted by noncircular interferences-application to SAIC," In *IEEE Transactions on Signal Processing*, Vol. 54, No. 3, pp. 870–883, Mar 2006.

[37] P. Chevalier, J.-P. Delmas, and A. Oukaci, "Optimal widely linear MVDR beamforming for noncircular signals," In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, Apr 2009, pp. 3573–3576.

[38] P. Chevalier and A. Blin, "Widely Linear MVDR Beamformers for the Reception of an Unknown Signal Corrupted by Noncircular Interferences," In *IEEE Transactions*

on Signal Processing, Vol. 55, No. 11, pp. 5323–5336, Nov 2007.

[39] G. Johnson, "The Complex (and Circular) Argument Continues," In *IEEE Signal Processing Magazine*, Vol. 13, No. 5, pp. 42–, Sept 1996.

[40] D. Lerro and Y. Bar-Shalom, "Tracking with debiased consistent converted measurements versus EKF," In *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 29, No. 3, pp. 1015–1022, Jul 1993.

[41] M. Longbin, S. Xiaoquan, Z. Yiyu, S. Z. Kang, and Y. Bar-Shalom, "Unbiased converted measurements for tracking," In *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 34, No. 3, pp. 1023–1027, Jul 1998.

[42] Y. C. Eldar, "Minimax estimation of deterministic parameters in linear models with a random model matrix," In *IEEE Transactions on Signal Processing*, Vol. 54, No. 2, pp. 601–612, Feb 2006.

[43] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," In *IEEE Journal on Selected Areas in Communications*, Vol. 21, No. 3, pp. 281–302, Apr 2003.

[44] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," In *IEEE Transactions on Signal Processing*, Vol. 51, No. 9, pp. 2257–2269, Sept 2003.

[45] M. Huemer, C. Hofbauer, and J. B. Huber, "Non-Systematic Complex Number RS Coded OFDM by Unique Word Prefix," In *IEEE Transactions on Signal Processing*, Vol. 60, No. 1, pp. 285–299, Jan 2012.

[46] A. Wiesel, Y. C. Eldar, and A. Yeredor, "Linear Regression With Gaussian Model Uncertainty: Algorithms and Bounds," In *IEEE Transactions on Signal Processing*, Vol. 56, No. 6, pp. 2194–2205, Jun 2008.

[47] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," In *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1–38, 1977.

[48] J. B. Rosen, H. Park, and J. Glick, "Total Least Norm Formulation and Solution for Structured Problems," In *SIAM Journal on Matrix Analysis and Applications*, Vol. 17, No. 1, pp. 110–126, 1996.

[49] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems," In *IEEE Transactions on Signal Processing*, Vol. 55, No. 5, pp. 2226–2238, May 2007.

[50] S. Ozen, C. Pladdy, M. J. Fimoff, S. M. Nerayanuru, and M. D. Zoltowski, "Approximate best linear unbiased channel estimation for frequency selective multipath

channels with long delay spreads," In *Proc. of the 37th Asilomar Conference on Signals, Systems and Computers*, Vol. 1, Nov 2003, pp. 1122–1127.

[51] M. Ghogho and A. Swami, "Improved channel estimation using superimposed training," In *Signal Processing Advances in Wireless Communications, 2004 IEEE 5th Workshop on*, Jul 2004, pp. 110–114.

[52] L. Lista, "The bias of the unbiased estimator: A study of the iterative application of the BLUE method," In *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 764, pp. 82 – 93, 2014.

[53] G. Box and G. Tiao, *Bayesian inference in statistical analysis*, ser. Addison-Wesley series in behavioral science: quantitative methods. Addison-Wesley Pub. Co., 1973.

[54] M. A. Woodbury, *Inverting Modified Matrices*, ser. Statistical Research Group Memorandum Reports. Princeton, NJ: Princeton University, 1950, No. 42.

[55] L. Fu-Mueller, M. Lunglmayr, and M. Huemer, "Reducing the gap between linear biased classical and linear Bayesian estimation," In *Proc. of the IEEE Statistical Signal Processing Workshop (SSP)*, Aug 2012, pp. 65–68.

[56] Y. C. Eldar, "Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound," In *Found. Trends Signal Process.*, Vol. 1, No. 4, pp. 305–449, April 2008.

[57] S. M. Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, ser. Fundamentals of Statistical Signal Processing. Prentice-Hall PTR, 2013.

[58] O. Lang, M. Huemer, and C. Hofbauer, "On the log-likelihood ratio evaluation of CWCU linear and widely linear MMSE data estimators," In *Proc. of the 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 633–637.

[59] M. Huemer and O. Lang, "On component-wise conditionally unbiased linear Bayesian estimation," In *Proc. of the 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 879–885.

[60] O. Lang and M. Huemer, "CWCU LMMSE Estimation Under Linear Model Assumptions," In *15th International Conference on Computer Aided Systems Theory – EUROCAST 2015, Las Palmas de Gran Canaria, Spain, February 8-13, 2015, Revised Selected Papers.* Springer International Publishing, 2015, pp. 537–545.

[61] M. Huemer, O. Lang, and C. Hofbauer, "Component-wise conditionally unbiased widely linear MMSE estimation," In *Signal Processing*, Vol. 133, pp. 227 – 239, 2017.

[62] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB.* John Wiley & Sons, 2010.

[63] W. Haselmayr, O. Lang, A. Springer, and M. Huemer, "Does Vector Gaussian Approximation After LMMSE Filtering Improve the LLR Quality?" In *IEEE Signal Processing Letters*, Vol. 24, No. 11, pp. 1676–1680, Nov 2017.

[64] S. Allpress, C. Luschi, and S. Felix, "Exact and approximated expressions of the log-likelihood ratio for 16-QAM signals," In *Proc. of the 38th Asilomar Conference on Signals, Systems and Computers*, Vol. 1. IEEE, 2004, pp. 794–798.

[65] IEEE Std 802.11a-1999, *Part11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-Speed Physical Layer in the 5 GHz Band*, 1999.

[66] M. Huemer, C. Hofbauer, A. Onic, and J. B. Huber, "Design and analysis of UW-OFDM signals," In *AEU-International Journal of Electronics and Communications*, Vol. 68, No. 10, pp. 958–968, 2014.

[67] J. Fakatselis, "Criteria for 2.4 GHz PHY comparison of modulation methods," In *Document IEEE; P802.11-97/157r1*, 1997.

[68] K. S. Kim, K. Hyun, C. W. Yu, Y. O. Park, D. Yoon, and S. K. Park, "General log-likelihood ratio expression and its implementation algorithm for Gray-Coded QAM signals," In *ETRI journal*, Vol. 28, No. 3, pp. 291–300, 2006.

[69] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," In *IEEE Transactions on Communications*, Vol. 50, No. 7, pp. 1074–1080, 2002.

[70] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," In *European Transactions on Telecommunications*, Vol. 8, No. 2, pp. 119–125, 1997.

[71] A. Sayed, *Fundamentals of Adaptive Filtering*, ser. Wiley - IEEE Series. Wiley, 2003.

[72] D. Mandic and V. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, ser. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2009.

[73] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," In *Proceedings of the IEEE*, Vol. 63, No. 4, pp. 719–720, Apr 1975.

[74] A. Gebhard, R. S. Kanumalli, B. Neurauter, and M. Huemer, "Adaptive self-interference cancelation in LTE-A carrier aggregation FDD direct-conversion transceivers," In *IEEE Sensor Array and Multichannel Signal Processing Workshop*

(SAM), Jul 2016, pp. 1–5.

[75] R. S. Kanumalli, A. Gebhard, A. Elmaghraby, A. Mayer, D. Schwartz, and M. Huemer, “Active Digital Cancellation of Transmitter Induced Modulated Spur Interference in 4G LTE Carrier Aggregation Transceivers,” In *Proc. of the IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.

[76] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, “Splitting the unit delay [FIR/all pass filters design],” In *IEEE Signal Processing Magazine*, Vol. 13, No. 1, pp. 30–60, Jan 1996.

[77] M. Lunglmayr, O. Lang, and M. Huemer, “Knowledge-Aided Kaczmarz and LMS Algorithms,” In *ArXiv e-prints arXiv:1712.02146 [eess.SP]*, December 2017.

[78] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* Wiley-Interscience, 2006.

[79] M. Lunglmayr, C. Unterrieder, and M. Huemer, “Approximate least squares,” In *Prof. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, 2014, pp. 4678–4682.

[80] S. Kaczmarz, “Przyblizone rozwiazywanie ukladów równan liniowych. - Angenaeherte Aufloesung von Systemen linearer Gleichungen.” In *Bulletin International de l’Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques,*, pp. 355–357, 1937.

[81] T. Tao, *Topics in random matrix theory.* American Mathematical Society, 1997.

[82] G. Deng, “Partial update and sparse adaptive filters,” In *IET Signal Processing*, Vol. 1, pp. 9–17(8), Mar 2007.

[83] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, “Compressed sensing Block MAP-LMS adaptive filter for sparse channel estimation and a Bayesian Cramer-Rao bound,” In *2009 IEEE International Workshop on Machine Learning for Signal Processing*, Sept 2009, pp. 1–6.

[84] A. Ma and D. Needell, “Adapted Stochastic Gradient Descent for Linear Systems with Missing Data,” In *arXiv:1702.07098 [math.NA]*, 2017. [Online]. Available: https://arxiv.org/abs/1702.07098

[85] S. Kay, *Intuitive Probability and Random Processes Using MATLAB.* Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.