COMBINING LINEAR ESTIMATION WITH SCALAR WIDELY LINEAR ESTIMATION

Christoph Hellings and Wolfgang Utschick

Professur für Methoden der Signalverarbeitung, Technische Universität München, Munich, Germany
Telephone: +49 89 289-28516, Fax: +49 89 289-28522, Email: {hellings,utschick}@tum.de

ABSTRACT

We study the capabilities of a filter configuration where a linear filter is applied to an improper vector, and the output is processed by scalar widely linear filters afterwards. Assuming that this filter configuration is used to estimate a vector of interest from a noisy observation, we aim at finding the optimal filter coefficients in the sense of minimizing the mean square error. To this end, we propose a filter design algorithm based on alternating optimization. The resulting filter turns out to achieve an intermediate performance between the optimal widely linear filter and the optimal linear filter. As an application example, we discuss how the considered filter structure fits to the concept of linear transceivers in communication systems.

Index Terms— improper signals, MMSE filter, widely linear.

1. INTRODUCTION

Complex signals are called improper if they have (linear) correlations with their own complex conjugates [1,2]. This is the case, e.g., for noncircular Gaussian signals. Whenever the observation and/or the quantity of interest in an estimation problem are improper, so-called widely linear estimation [3] can outperform conventional linear estimators. On the other hand, widely linear estimation does not bring any benefits if all signals in a system are proper.

When implementing widely linear filtering via a real-valued representation (e.g. [2]), the computational complexity of the filtering operations is generally not higher than for linear filters. However, it can still be preferable to stick to linear filters in some applications, e.g., because a lower number of filter coefficients need to be stored, or if the aim is to improve an existing system that is based on linear filters without drastic changes to the structure of the system. Another example are communication systems, where widely linear filtering can be easily implemented before the modulation stage and after the demodulation stage, but not for a modulated signal where the inphase and quadrature components are superimposed.

Our aim is to study whether knowledge about the impropriety of the involved signals can be exploited by introducing some aspects of widely linear filtering into the system design without completely replacing linear filters by widely linear filters. As a first step towards answering this question, this paper considers a simple estimation problem, where a random vector of interest is to be estimated from a noisy observation vector. In this setting, we study a two-stage filter that first applies a conventional linear filter to the whole vector and performs component-wise widely linear filtering afterwards.

A two-stage approach for widely linear filtering was also considered in [4], but that method estimates scalars (instead of vectors) by first performing a dimensionality reduction and then using a general widely linear filter in the low-dimensional space (instead of applying a component-wise widely linear filter with full dimensional). In [5], the coexistence of improper signals and linear transmit filters was studied, but the question we consider here did not arise since no restriction on the receive filters (estimators) were imposed in [5].

In Section 2, we give motivations for this particular filter structure, and we briefly review the theory of improper signals and widely linear operations. We then propose a simple filter design method that already brings gains over purely linear filtering (Section 3), and we further improve the filter by an iterative optimization (Section 4). After some numerical experiments in Section 5, we give an outlook to related questions for future research in Section 6.

Notation: For easy distinction, we use sans-serif font for complex quantities, and serif font for real quantities.

2. MOTIVATION AND MATHEMATICAL BACKGROUND

Consider the estimation of a scalar real-valued random variable x from a scalar complex observation $y = hx + n$, where $h \in \mathbb{C}$, and n is circularly symmetric complex Gaussian noise. Figure 1 illustrates an example where x comes from a finite set $\{\pm d\}$. The estimator should rotate the signal of interest back to the real axis and adjust the magnitude in a way that the mean square error (MSE) is minimized.

It obviously makes sense to apply a real part operator in addition since we know that the imaginary part contains only noise after an appropriate rotation. However, this last step cannot be performed by a linear filter: the real part is not a linear operation since $\Re(ax) = a\Re(x)$ is not fulfilled for all $a \in \mathbb{C}$. Instead, it is a very simple special case of a so-called widely linear operation.

Let us now extend this idea to vector-valued scenarios and to scenarios where we do not deal with real-valued signals, but with arbitrary improper random vectors. Since a simple intuitive argumentation as above is no longer possible, we need to introduce the concepts of improper random vectors and of widely linear filters. In addition to the conventional covariance matrix $\boldsymbol{C}_{\boldsymbol{x}} = \mathrm{E}[(\boldsymbol{x} - \mathrm{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathrm{E}[\boldsymbol{x}])^{\mathrm{H}}]$, we can define the pseudocovariance matrix [1]

$$\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \mathrm{E}[(\boldsymbol{x} - \mathrm{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathrm{E}[\boldsymbol{x}])^{\mathrm{T}}]. \tag{1}$$

Note that we assume mean zero for all random vectors throughout this paper for the sake of simplicity, so that we have $\boldsymbol{C}_{\boldsymbol{x}} = \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}]$ and $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]$. If the pseudocovariance matrix vanishes, the random vector is called proper [1]. Otherwise, i.e., if $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} \neq \mathbf{0}$, the vector $\boldsymbol{x}$ is called improper [2]. In the above example, it is easy to verify that both scalars x and y are improper.

A widely linear operation $\boldsymbol{y} \mapsto \boldsymbol{f}(\boldsymbol{y}) = \boldsymbol{A}_{\mathrm{L}}\boldsymbol{y} + \boldsymbol{A}_{\mathrm{CL}}\boldsymbol{y}^*$ [3] can be equivalently written as

$$\boldsymbol{f}(\boldsymbol{y}) = (\boldsymbol{A}_{\mathrm{L}} + \boldsymbol{A}_{\mathrm{CL}})\Re(\boldsymbol{y}) + \mathrm{j}(\boldsymbol{A}_{\mathrm{L}} - \boldsymbol{A}_{\mathrm{CL}})\Im(\boldsymbol{y}) \tag{2}$$

i.e., it can be implemented by first splitting the input into its real and imaginary parts, and then performing individual linear filtering of the two components. However, first performing a linear filtering of

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

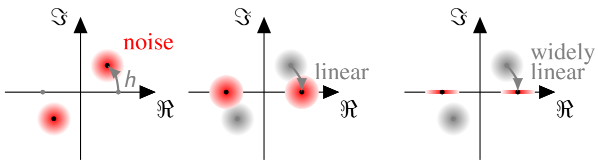

Fig. 1: In addition to compensating the effect of h, a widely linear filter can remove a real noise dimension in this example.

the complete vector and applying real and imaginary part operators afterwards—in analogy to the minimal example from above—is in general not an equivalent structure. To see this, note that the number of filter coefficients in the two matrices $\boldsymbol{A}_{\mathrm{L}}$ and $\boldsymbol{A}_{\mathrm{CL}}$ is twice as high as in a linear filter, where the conjugate linear part $\boldsymbol{A}_{\mathrm{CL}}$ is zero.

Consequently, the full potential of widely linear filtering can in general not be exploited by keeping a linear filter and applying scalar widely linear operations to the components of the filter output. The question is thus how far we can get with the restrictive filter structure

$$\hat{\boldsymbol{x}} = \operatorname{diag}\{a_k\}\boldsymbol{z} + \operatorname{diag}\{b_k\}\boldsymbol{z}^* \quad \text{with} \quad \boldsymbol{z} = \boldsymbol{G}\boldsymbol{y} \tag{3}$$

where diag is used to denote a diagonal matrix. Since $a_k, b_k \in \mathbb{C}$ for $k = 1, \ldots, K$, and $\boldsymbol{G} \in \mathbb{C}^{K \times N}$, this description has a total of $KN + 2K$ complex coefficients compared to $2KN$ in case of a general widely linear filter.[1]

2.1. Application Example: Linear Transceivers

As a second motivation, consider a communication scenario where data is transmitted by means of improper signaling. This can help to reduce the harmfulness of the interference to other users, and was proposed, e.g., for secondary users in cognitive radio systems [6, 7].

A common assumption are so-called linear transceivers [8]. In this case, the data transmission is described by $\hat{\boldsymbol{x}} = \boldsymbol{R}(\boldsymbol{H}\boldsymbol{T}\boldsymbol{x} + \boldsymbol{\eta})$, where $\boldsymbol{T}$ and $\boldsymbol{R}$ are linear transmit and receive filters, $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ is the channel matrix, and $\boldsymbol{\eta}$ is additive Gaussian noise and interference. In the context of linear transceivers, the symbol vector $\boldsymbol{x}$ is usually assumed to be circularly symmetric (and thus proper) Gaussian with mean zero and $\boldsymbol{C}_{\boldsymbol{x}} = \mathbf{I}_K$, where K is the number of data streams.

If the aim is to transmit improper signals instead, it was proposed to replace the linear transceivers by widely linear transceivers [5, 9], i.e., to replace the linear transmit and receive filters by widely linear filters. We are now interested in the question whether transmission of improper signals can also be performed in a sensible manner with linear transceivers. For simplicity, we assume $K = M = N$.

Since digital transmission always incorporates nonlinear operations, such as encoding, detection, etc., linear transceivers are usually understood as transmit strategies where nonlinear operations are only applied to single data streams while all filtering operations that involve multiple data streams have to be linear [9]. According to this definition, applying a widely linear filter to a single stream directly after encoding or directly before the detection does not violate the assumption of linear transceivers. On the other hand, if a widely linear operation is applied to a whole vector of data streams, we have implemented widely linear transceivers instead [5, 9].

Let us first consider the transmitter side. The achievable rate over the link is related to the mutual information between the transmit signal $\boldsymbol{s} = \boldsymbol{T}\boldsymbol{x}$ and the received signal $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{T}\boldsymbol{x} + \boldsymbol{\eta}$. Since all signals are assumed to be Gaussian, this mutual information depends only on the second-order properties of the transmit signal and of the noise, i.e., on their covariance matrices and pseudocovariance matrices (see, e.g., [10]). For any pair of $\boldsymbol{C}_{\boldsymbol{s}}$ and $\tilde{\boldsymbol{C}}_{\boldsymbol{s}}$ that we want to create, we can consider a special form of the singular value decomposition

$$\boldsymbol{U}\boldsymbol{K}\boldsymbol{V}^{\mathrm{H}} = \boldsymbol{C}_{\boldsymbol{s}}^{-\frac{1}{2}}\tilde{\boldsymbol{C}}_{\boldsymbol{s}}\boldsymbol{C}_{\boldsymbol{s}}^{-\frac{\mathrm{T}}{2}} \quad \text{with} \quad \boldsymbol{V}^{\mathrm{H}} = \boldsymbol{U}^{\mathrm{T}} \tag{4}$$

which is called Takagi's factorization and exists since $\boldsymbol{C}_{\boldsymbol{s}}^{-\frac{1}{2}}\tilde{\boldsymbol{C}}_{\boldsymbol{s}}\boldsymbol{C}_{\boldsymbol{s}}^{-\frac{\mathrm{T}}{2}}$ is a complex symmetric matrix [2]. The diagonal entries of $\boldsymbol{K}$ can be interpreted as impropriety coefficients [2], which indicate the strength of impropriety. To create a transmit vector with the desired second-order properties, we can generate a vector $\boldsymbol{x}$ with covariance matrix $\boldsymbol{C}_{\boldsymbol{x}} = \mathbf{I}_K$ and pseudocovariance matrix $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \boldsymbol{K}$ and process this vector by a strictly linear filter $\boldsymbol{T} = \boldsymbol{C}_{\boldsymbol{s}}^{\frac{1}{2}}\boldsymbol{U}$. It is easy to verify that this indeed leads to $\mathrm{E}[\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}] = \boldsymbol{C}_{\boldsymbol{s}}$ and $\mathrm{E}[\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}] = \tilde{\boldsymbol{C}}_{\boldsymbol{s}}$.

As $\boldsymbol{C}_{\boldsymbol{x}}$ and $\tilde{\boldsymbol{C}}_{\boldsymbol{x}}$ are both diagonal, $\boldsymbol{x}$ has uncorrelated entries, and the individual impropriety of the components can be created already as a part of the nonlinear encoding and symbol mapping. Thus, arbitrary improper signals can be transmitted with a linear transmitter.[2]

We now turn our attention to the receiver side. Under ideal conditions, namely if the channel can be inverted and the noise is negligible, the linear filter $\boldsymbol{R} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{s}}^{-\frac{1}{2}}\boldsymbol{H}^{-1}$ can recover the vector $\boldsymbol{x}$. As the components of the recovered vector are uncorrelated, the nonlinear detection and decoding operations can operate on each component separately and can account for the individual impropriety of the components. In this paper, we restrict our considerations to an MSE perspective and do not focus on the further nonlinear operations. Therefore, we process the components by scalar widely linear filters to minimize the MSE. This does not contradict the assumption of linear transceivers as it happens on a per-stream basis.

However, if we take the noise into account, the above receive filter leads to correlated noise, making per-stream decoding suboptimal. On the other hand, transceivers that are designed such that correlations of the errors on the individual streams are avoided cannot guarantee that cross-improprieties between the streams are removed. Due to these conflicting interests, linear transceivers can in general not achieve the same performance as widely linear transceivers if improper signals are involved. In this paper, we study a best-effort approach for dealing with impropriety as good as possible (in terms of minimal MSE) despite the restrictions to the transceiver structure.

2.2. Minimum Mean Square Error Criterion

Our design criterion in the noisy case is the minimum mean square error (MMSE) criterion, i.e., we aim at solving

$$\min_{\boldsymbol{G},\ (a_k, b_k)_{\forall k}} \mathrm{E}[\|\operatorname{diag}\{a_k\}\boldsymbol{G}\boldsymbol{y} + \operatorname{diag}\{b_k\}(\boldsymbol{G}\boldsymbol{y})^* - \boldsymbol{x}\|_2^2]. \tag{5}$$

As it is not obvious how to solve this problem jointly in both groups of variables, we resort to an alternating optimization approach. Although we do not claim this method to be globally optimal (see Section 6), it leads to remarkable results in numerical simulations.

In order to not limit our considerations to the application example described above, the derivations in the following sections apply to a general estimation problem where a quantity of interest $\boldsymbol{x}$ has to be estimated from a noisy observation $\boldsymbol{y}$, and the connection between

[1] This representation is not unique since one of the factors a_k and b_k for each k could be moved into the matrix $\boldsymbol{G}$. However, this ambiguous formulation facilitates the following considerations. For an implementation, the number of coefficients could then be reduced to $(N + 1)K$.

[2] An alternative would be to start with a proper $\boldsymbol{x}$ and to introduce the component-wise impropriety by means of scalar widely linear filters. The combination of these filters with the the linear transmit filter would then be exactly the reverse structure of the filter structure considered in this paper.

these two quantities is described by a given cross-covariance matrix $\boldsymbol{C}_{\boldsymbol{xy}} = \mathrm{E}[\boldsymbol{x}\boldsymbol{y}^{\mathrm{H}}]$ and a cross-pseudocovariance matrix $\tilde{\boldsymbol{C}}_{\boldsymbol{xy}} = \mathrm{E}[\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}]$. Moreover, the individual second-order properties of $\boldsymbol{x}$ and $\boldsymbol{y}$ are described by the covariance matrices $\boldsymbol{C}_{\boldsymbol{x}}$, $\boldsymbol{C}_{\boldsymbol{y}}$ and by the pseudocovariance matrices $\tilde{\boldsymbol{C}}_{\boldsymbol{x}}$, $\tilde{\boldsymbol{C}}_{\boldsymbol{y}}$. For the data transmission example, we have

$$\boldsymbol{C}_{\boldsymbol{x}} = \mathbf{I}_K \qquad \tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \boldsymbol{K} \tag{6}$$

$$\boldsymbol{C}_{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{\eta}} \qquad \tilde{\boldsymbol{C}}_{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{T}\boldsymbol{K}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}} + \tilde{\boldsymbol{C}}_{\boldsymbol{\eta}} \tag{7}$$

$$\boldsymbol{C}_{\boldsymbol{xy}} = \boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}} \qquad \tilde{\boldsymbol{C}}_{\boldsymbol{xy}} = \boldsymbol{K}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}. \tag{8}$$

3. BASIC FILTER DESIGN

Before introducing the iterative alternating optimization method, we first study a simple heuristic approach. To this end, we design the linear filter $\boldsymbol{G}$ as if it was the only filter, i.e., we optimize $\boldsymbol{G}$ for fixed $a_k = 1$ and $b_k = 0$ for all k. The optimization can be rewritten as

$$\min_{\boldsymbol{G}} \mathrm{E}[\|\boldsymbol{G}\boldsymbol{y} - \boldsymbol{x}\|_2^2] \tag{9}$$

which is solved by a conventional linear MMSE filter

$$\boldsymbol{G} = \boldsymbol{C}_{\boldsymbol{xy}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}. \tag{10}$$

In a second step, we fix $\boldsymbol{G}$ and optimize over a_k and b_k, i.e.,

$$\min_{(a_k,b_k)\forall k} \sum_{k=1}^{K} \mathrm{E}[|a_k z_k + b_k z_k^* - x_k|^2] \tag{11}$$

with $\boldsymbol{z} = [z_1, \ldots, z_K]^{\mathrm{T}} = \boldsymbol{G}\boldsymbol{y}$. Obviously, this optimization can be solved for every component separately, and the solution is given by the optimal widely linear MMSE filter [3]

$$a_k = \frac{c_{x_k z_k} - \tilde{c}_{x_k z_k} c_{z_k}^{-1} \tilde{c}_{z_k}^*}{c_{z_k} - |\tilde{c}_{z_k}|^2 c_{z_k}^{-1}} \qquad b_k = \frac{\tilde{c}_{x_k z_k} - c_{x_k z_k} c_{z_k}^{-1} \tilde{c}_{z_k}}{c_{z_k} - |\tilde{c}_{z_k}|^2 c_{z_k}^{-1}} \tag{12}$$

for estimating x_k from z_k. Here,

$$c_{z_k} = \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{G}\boldsymbol{C}_{\boldsymbol{y}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{e}_k \qquad \tilde{c}_{z_k} = \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{G}\tilde{\boldsymbol{C}}_{\boldsymbol{y}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{e}_k \tag{13}$$

$$c_{x_k z_k} = \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{xy}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{e}_k \qquad \tilde{c}_{x_k z_k} = \boldsymbol{e}_k^{\mathrm{T}}\tilde{\boldsymbol{C}}_{\boldsymbol{xy}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{e}_k. \tag{14}$$

where $\boldsymbol{e}_k$ is the kth canonical unit vector.

From [3], it is known that the optimal widely linear MMSE filter strictly outperforms the optimal linear MMSE filter unless

$$\tilde{c}_{x_k z_k} - c_{x_k z_k} c_{z_k}^{-1} \tilde{c}_{z_k}^* = 0 \tag{15}$$

in which case both filters perform equally. In particular, (15) holds if all involved signals are jointly proper. On the other hand, if improper signals are involved, the condition is in general not fulfilled.

Since $\boldsymbol{G}$ is the optimal linear MMSE filter, the original choice $a_k = 1$ and $b_k = 0$ (where the scalar signal is not modified any further) corresponds to the optimal scalar linear MMSE filter. Replacing this choice by the widely linear solution (12) leads to a strict reduction of the mean square error, except in cases where (15) holds.

4. ITERATIVE FILTER DESIGN

After optimizing the coefficients a_k and b_k of the scalar widely linear filters, the assumptions under which $\boldsymbol{G}$ had been optimized ($a_k = 1$ and $b_k = 0$) are no longer valid. This means that we can potentially find an improved linear filter $\boldsymbol{G}$ by solving

$$\min_{\boldsymbol{G}} \mathrm{E}[\|\operatorname{diag}\{a_k\}\boldsymbol{G}\boldsymbol{y} + \operatorname{diag}\{b_k\}(\boldsymbol{G}\boldsymbol{y})^* - \boldsymbol{x}\|_2^2] \tag{16}$$

after plugging in the solutions for a_k and b_k from (12). If this leads to a different solution for $\boldsymbol{G}$, we can again update a_k and b_k, and we can repeat this procedure until convergence. Since solving the optimization in (11) or (16) cannot increase the MSE, this iteration produces a monotonic sequence of MSE values, so that convergence in terms of the MSE is guaranteed.

The remaining task is to derive a solution to (16). The qualitative difference between this problem and the optimization in (9) is that we are now looking for a linear filter that minimizes the MSE of an expression that involves a subsequently applied widely linear filter. Therefore, the well-known equation (10) can no longer be used.

4.1. Composite Real Representation

We use the composite real representations

$$\check{\boldsymbol{v}} = \begin{bmatrix} \Re(\boldsymbol{v}) \\ \Im(\boldsymbol{v}) \end{bmatrix} \qquad \grave{\boldsymbol{M}} = \begin{bmatrix} \Re(\boldsymbol{M}) & -\Im(\boldsymbol{M}) \\ \Im(\boldsymbol{M}) & \Re(\boldsymbol{M}) \end{bmatrix} \tag{17}$$

of complex vectors and matrices to simplify the following derivations. A widely linear transformation (2) can then be written as (e.g., [11, Th. 2])

$$\check{\boldsymbol{f}}(\check{\boldsymbol{y}}) = \boldsymbol{A}\check{\boldsymbol{y}} \quad \text{with} \quad \boldsymbol{A} = \grave{\boldsymbol{A}}_{\mathrm{L}} + \acute{\boldsymbol{A}}_{\mathrm{CL}} \tag{18}$$

where

$$\acute{\boldsymbol{A}}_{\mathrm{CL}} = \begin{bmatrix} \Re(\boldsymbol{A}_{\mathrm{CL}}) & \Im(\boldsymbol{A}_{\mathrm{CL}}) \\ \Im(\boldsymbol{A}_{\mathrm{CL}}) & -\Re(\boldsymbol{A}_{\mathrm{CL}}) \end{bmatrix}. \tag{19}$$

Matrices with the block structures in (17) and (19) form two orthogonal subspaces which together span the whole vector space of real-valued matrices of appropriate dimensions [11, Lemma 5].

This implies that a real-valued linear mapping $\check{\boldsymbol{y}} \mapsto \boldsymbol{A}\check{\boldsymbol{y}}$ corresponds to a linear mapping in the complex domain only if $\boldsymbol{A}$ has the special block structure shown in (17). In [11, 12], such matrices were called block-skew-circulant with 2×2 blocks ($\mathcal{BSC}_2$), and they were called matrices with complex structure in [13]. Matrices with the block structure in (19) were called block-Hankel-skew-circulant with 2×2 blocks ($\mathcal{BHSC}_2$) in [11]. In this paper, we use $\mathbb{L}$ to denote the subspace of $\mathcal{BSC}_2$ matrices (where the dimensions become clear from the context), and $\mathbb{L}^{\perp}$ for the orthogonal complement, i.e., for the subspace of $\mathcal{BHSC}_2$ matrices. We use grave and acute accents ($\grave{\bullet}$ and $\acute{\bullet}$) to denote $\mathcal{BSC}_2$ or $\mathcal{BHSC}_2$ matrices, respectively.

The orthogonal projections to $\mathbb{L}$ and $\mathbb{L}^{\perp}$ read as [11, Lemma 6]

$$\mathrm{P}_{\mathbb{L}}\left(\begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} \boldsymbol{C}_1 + \boldsymbol{C}_4 & \boldsymbol{C}_2 - \boldsymbol{C}_3 \\ \boldsymbol{C}_3 - \boldsymbol{C}_2 & \boldsymbol{C}_1 + \boldsymbol{C}_4 \end{bmatrix} \tag{20}$$

$$\mathrm{P}_{\mathbb{L}^{\perp}}\left(\begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} \boldsymbol{C}_1 - \boldsymbol{C}_4 & \boldsymbol{C}_2 + \boldsymbol{C}_3 \\ \boldsymbol{C}_2 + \boldsymbol{C}_3 & \boldsymbol{C}_4 - \boldsymbol{C}_1 \end{bmatrix} \tag{21}$$

and we note the following result from [11, Lemma 3].

Lemma 1. *Let $\grave{\boldsymbol{A}}$, $\grave{\boldsymbol{A}}'$ be $\mathcal{BSC}_2$ and $\acute{\boldsymbol{B}}$, $\acute{\boldsymbol{B}}'$ be $\mathcal{BHSC}_2$ with appropriate sizes. Then, $\grave{\boldsymbol{A}}\grave{\boldsymbol{A}}'$ is $\mathcal{BSC}_2$, $\acute{\boldsymbol{B}}\acute{\boldsymbol{B}}'$ is $\mathcal{BSC}_2$, $\grave{\boldsymbol{A}}\acute{\boldsymbol{B}}'$ is $\mathcal{BHSC}_2$, and $\acute{\boldsymbol{B}}\grave{\boldsymbol{A}}'$ is $\mathcal{BHSC}_2$.*

4.2. Optimal Linear Filter

In the real-valued representation, we can write (16) as

$$\min_{\grave{\boldsymbol{G}}} \mathrm{E}[\|\boldsymbol{S}\grave{\boldsymbol{G}}\check{\boldsymbol{y}} - \check{\boldsymbol{x}}\|_2^2] \quad \text{s.t.} \quad \mathrm{P}_{\mathbb{L}^{\perp}}(\grave{\boldsymbol{G}}) = \boldsymbol{0} \tag{22}$$

where $\boldsymbol{S} = \grave{\boldsymbol{S}}_{\mathrm{L}} + \acute{\boldsymbol{S}}_{\mathrm{CL}}$ with $\boldsymbol{S}_{\mathrm{L}} = \operatorname{diag}\{a_k\}$ and $\boldsymbol{S}_{\mathrm{CL}} = \operatorname{diag}\{b_k\}$. As $\grave{\boldsymbol{G}}$ is constrained to be a $\mathcal{BSC}_2$ matrix, it corresponds to a complex linear operation. After solving (22), the optimizer $\boldsymbol{G}$ of (16) for given a_k and b_k can thus be obtained via (17).

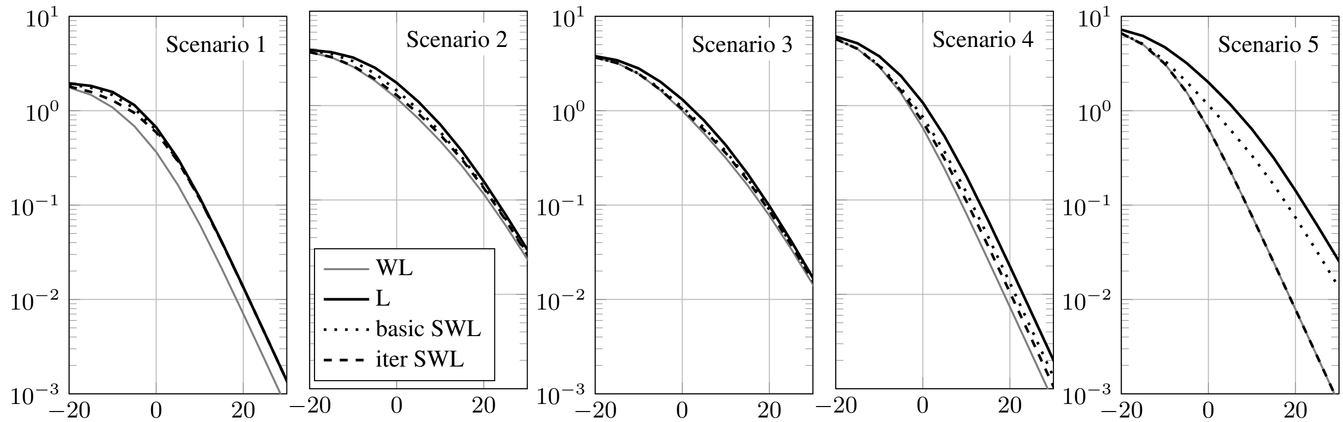

Fig. 2: Achieved MSE over SNR σ^{-2} in dB. Scenario 1: $M = 2$, $N = 4$, $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \mathbf{0}$, $\tilde{\boldsymbol{C}}_{\boldsymbol{\eta}} = \sigma^2 0.9 \mathbf{I}_4$. Scenario 2: $M = N = 4$, $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \operatorname{diag}\{0.9, 0.8, 0.7, 0.6\}$, $\tilde{\boldsymbol{C}}_{\boldsymbol{\eta}} = \sigma^2 \operatorname{diag}\{0.9, 0.8, 0.7, 0.6\}$. Scenario 3: $M = N = 4$, $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \operatorname{diag}\{0.9, 0.8, 0.7, 0.6\}$, $\tilde{\boldsymbol{C}}_{\boldsymbol{\eta}} = \mathbf{0}$. Scenario 4: $M = 6$, $N = 8$, $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \operatorname{diag}\{1, 1, 1, 1, 0, 0\}$, $\tilde{\boldsymbol{C}}_{\boldsymbol{\eta}} = \mathbf{0}$. Scenario 5: $M = N = 8$, $\tilde{\boldsymbol{C}}_{\boldsymbol{x}} = \mathbf{I}_8$, $\tilde{\boldsymbol{C}}_{\boldsymbol{\eta}} = \mathbf{0}$.

We introduce a Lagrangian multiplier $\boldsymbol{\Lambda} = \grave{\boldsymbol{\Lambda}} + \acute{\boldsymbol{\Lambda}}$, and we use the trace operator $\operatorname{tr}[\bullet]$ to write the Lagrangian function

$$\begin{aligned} \Phi = \operatorname{tr}[\boldsymbol{S}\grave{\boldsymbol{G}}\boldsymbol{C}_{\check{\boldsymbol{y}}}\grave{\boldsymbol{G}}^{\mathrm{T}}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{S}\grave{\boldsymbol{G}}\boldsymbol{C}_{\check{\boldsymbol{x}}\check{\boldsymbol{y}}}^{\mathrm{T}} + \boldsymbol{C}_{\check{\boldsymbol{x}}\check{\boldsymbol{y}}}\grave{\boldsymbol{G}}^{\mathrm{T}}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{C}_{\check{\boldsymbol{x}}}] \\ + \operatorname{tr}[\boldsymbol{\Lambda}^{\mathrm{T}}\, \mathrm{P}_{\mathbb{L}^\perp}(\grave{\boldsymbol{G}})] \end{aligned} \quad (23)$$

where $\boldsymbol{C}_{\check{\boldsymbol{y}}} = \mathrm{E}[\check{\boldsymbol{y}}\check{\boldsymbol{y}}^{\mathrm{T}}]$, $\boldsymbol{C}_{\check{\boldsymbol{x}}} = \mathrm{E}[\check{\boldsymbol{x}}\check{\boldsymbol{x}}^{\mathrm{T}}]$, and $\boldsymbol{C}_{\check{\boldsymbol{x}}\check{\boldsymbol{y}}} = \mathrm{E}[\check{\boldsymbol{x}}\check{\boldsymbol{y}}^{\mathrm{T}}]$. As the derivative of the last summand in (23) is $\mathrm{P}_{\mathbb{L}^\perp}(\boldsymbol{\Lambda}) = \acute{\boldsymbol{\Lambda}}$ due to [12, Appendix A], we obtain

$$\frac{\partial \Phi}{\partial \grave{\boldsymbol{G}}} = 2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}\grave{\boldsymbol{G}}\boldsymbol{C}_{\check{\boldsymbol{y}}} - 2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}_{\check{\boldsymbol{x}}\check{\boldsymbol{y}}} + \acute{\boldsymbol{\Lambda}}. \quad (24)$$

To find a pair $\acute{\boldsymbol{\Lambda}} \in \mathbb{L}^\perp$, $\grave{\boldsymbol{G}} \in \mathbb{L}$ such that $\frac{\partial \Phi}{\partial \grave{\boldsymbol{G}}} = \mathbf{0}$, we define

$$\grave{\boldsymbol{D}} + \acute{\boldsymbol{D}} = \boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}, \quad \grave{\boldsymbol{E}} + \acute{\boldsymbol{E}} = \boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}_{\check{\boldsymbol{x}}\check{\boldsymbol{y}}}, \quad \grave{\boldsymbol{F}} + \acute{\boldsymbol{F}} = \boldsymbol{C}_{\check{\boldsymbol{y}}} \quad (25)$$

where the separation into a $\mathcal{BSC}_2$ component and a $\mathcal{BHSC}_2$ component can be performed using the projections $\mathrm{P}_{\mathbb{L}}$ and $\mathrm{P}_{\mathbb{L}^\perp}$. Using Lemma 1, we can write down two separate equations for the $\mathcal{BSC}_2$ and the $\mathcal{BHSC}_2$ components of $\frac{\partial \Phi}{\partial \grave{\boldsymbol{G}}}$. We obtain

$$\mathbf{0} \overset{!}{=} 2(\grave{\boldsymbol{D}}\grave{\boldsymbol{G}}\grave{\boldsymbol{F}} + \acute{\boldsymbol{D}}\grave{\boldsymbol{G}}\acute{\boldsymbol{F}} - \grave{\boldsymbol{E}}) \in \mathbb{L} \quad (26)$$

$$\mathbf{0} \overset{!}{=} 2(\grave{\boldsymbol{D}}\grave{\boldsymbol{G}}\acute{\boldsymbol{F}} + \acute{\boldsymbol{D}}\grave{\boldsymbol{G}}\grave{\boldsymbol{F}} - \acute{\boldsymbol{E}}) + \acute{\boldsymbol{\Lambda}} \in \mathbb{L}^\perp \quad (27)$$

where we have exploited that $\grave{\boldsymbol{G}}$ is a $\mathcal{BSC}_2$ matrix due to the constraint $\grave{\boldsymbol{G}} \in \mathbb{L}$. Since $\acute{\boldsymbol{\Lambda}}$ is an arbitrary $\mathcal{BHSC}_2$ matrix, it is always possible to choose $\acute{\boldsymbol{\Lambda}}$ such that (27) is fulfilled. Using $\operatorname{vec}(\bullet)$ to denote the vectorization of a matrix and $\otimes$ for the Kronecker product, we can obtain the optimizer of (22) by solving (26) for

$$\operatorname{vec}(\grave{\boldsymbol{G}}) = \left((\grave{\boldsymbol{F}}^{\mathrm{T}} \otimes \grave{\boldsymbol{D}}) + (\acute{\boldsymbol{F}}^{\mathrm{T}} \otimes \acute{\boldsymbol{D}})\right)^{-1} \operatorname{vec}(\grave{\boldsymbol{E}}). \quad (28)$$

5. NUMERICAL RESULTS

For the numerical simulations, we use the example of a data transmission as introduced in Section 2.1. All results are averaged over 1000 realizations of the channel matrix $\boldsymbol{H}$, whose entries are generated as i.i.d. circularly symmetric Gaussian random variables with zero-mean and unit variance. For the sake of simplicity, the transmit filter is set to $\boldsymbol{T} = \mathbf{I}_M$, i.e., we have $K = M$. To facilitate the interpretation, we have chosen (scaled) identity covariance matrices $\boldsymbol{C}_{\boldsymbol{x}} = \mathbf{I}_M$ and $\boldsymbol{C}_{\boldsymbol{\eta}} = \sigma^2 \mathbf{I}_N$, and diagonal pseudocovariance matrices whose diagonal entries can (after normalization by σ^2) be directly interpreted as impropriety coefficients (see, e.g., [2]). Further simulation parameters are specified in the caption of Fig. 2.

In the first scenario, where we have a proper transmit signal, but improper noise, scalar widely linear (SWL) filtering can bring benefits over linear (L) filtering only in the low-SNR regime. For high SNR, widely linear (WL) filtering clearly outperforms the other approaches. An improvement due to SWL can be observed over a larger SNR range in Scenario 2 (improper signal and noise) and Scenario 3 (improper signal, but proper noise). Scenario 4 shows an example where $\boldsymbol{x}$ contains proper components and maximally improper components while the noise is proper. It is noteworthy that all curves converge to the same high-SNR slope here, but notable offsets between all schemes remain. The most remarkable result is observed in Scenario 5, where $\boldsymbol{x}$ is maximally improper. While purely linear filtering leads to a worse slope in this scenario, the SWL filter obtained by alternating optimization achieves the same performance as a widely linear filter. This effect occurs also with improper noise, but the plot for improper noise is omitted due to space constraints.

6. CONCLUSION AND OUTLOOK

Our simulations have revealed that the considered combination of linear estimation and scalar widely linear estimation can in some cases be an interesting alternative to a vector-valued widely linear filter. However, we have also observed that this approach does not always exploit the full potential of widely linear estimation.

It is currently an open question whether the employed alternating optimization method converges to the global optimum among all filters with the considered structure. This should be studied in future, e.g., based on the sufficient conditions from [14, 15].

Further possible research questions arise in the application for data transmission with linear transceivers. While the considerations in this paper are restricted to the MSE, it would be interesting to see in addition how achievable rates can be optimized in a system that transmits improper signals by means of (strictly) linear transceivers.

7. REFERENCES

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

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

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

[4] N. Song, W. U. Alokozai, R. C. de Lamare, and M. Haardt, "Adaptive widely linear reduced-rank beamforming based on joint iterative optimization," *IEEE Signal Process. Lett.*, vol. 21, no. 3, pp. 265–269, Mar. 2014.

[5] S. Lagen, A. Agustin, and J. Vidal, "Coexisting linear and widely linear transceivers in the MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 64, no. 3, pp. 652–664, Feb. 2016.

[6] C. Lameiro, I. Santamaría, and P. J. Schreier, "Benefits of improper signaling for underlay cognitive radio," *IEEE Wireless Commun. Lett.*, vol. 4, no. 1, pp. 22–25, Feb. 2015.

[7] O. Amin, W. Abediseid, and M. S. Alouini, "Underlay cognitive radio systems with improper Gaussian signaling: Outage performance analysis," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4875–4887, Jul. 2016.

[8] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends® Commun. and Inf. Theory*, vol. 3, no. 4–5, pp. 331–551, 2006.

[9] C. Hellings, M. Joham, and W. Utschick, "QoS feasibility in MIMO broadcast channels with widely linear transceivers," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1134–1137, Nov. 2013.

[10] Y. Zeng, C. M. Yetis, E. Gunawan, Y. L. Guan, and R. Zhang, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2899–2913, Jun. 2013.

[11] C. Hellings and W. Utschick, "Block-skew-circulant matrices in complex-valued signal processing," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2093–2107, Apr. 2015.

[12] C. Hellings, M. Koller, and W. Utschick, "An impropriety test based on block-skew-circulant matrices," presented at the Int. ITG Workshop on Smart Antennas (WSA) 2015, Ilmenau, Germany, Mar., 3–5 2015.

[13] S. A. Andersson, H. K. Brons, and S. T. Jensen, "Distribution of eigenvalues in multivariate statistical analysis," *Ann. Statist.*, vol. 11, no. 2, pp. 392–415, Jun. 1983.

[14] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions*, vol. Supplement Issue No. 1, pp. 205–237, 1984.

[15] U. Niesen, D. Shah, and G. Wornell, "Adaptive alternating minimization algorithms," in *Proc. Int. Symp. Inf. Theory (ISIT) 2007*, Jun. 2007, pp. 1641–1645.