Correspondence

Detection and Estimation of Improper Complex Random Signals

Peter J. Schreier, *Member, IEEE*, Louis L. Scharf, *Fellow, IEEE*, and Clifford T. Mullis

***Abstract*—Nonstationary complex random signals are in general improper (not circularly symmetric), which means that their complementary covariance is nonzero. Since the Karhunen–Loève (K-L) expansion in its known form is only valid for proper processes, we derive the improper version of this expansion. It produces two sets of eigenvalues and improper observable coordinates. We then use the K-L expansion to solve the problems of detection and estimation of improper complex random signals in additive white Gaussian noise. We derive a general result comparing the performance of conventional processing, which ignores complementary covariances, with processing that takes these into account. In particular, for the detection and estimation problems considered, we find that the performance gain, as measured by deflection and mean-squared error (MSE), respectively, can be as large as a factor of 2. In a communications example, we show how this finding generalizes the result that coherent processing enjoys a 3-dB gain over noncoherent processing.**

***Index Terms*—Detection, estimation, improper complex random signal, Karhunen–Loève (K-L) expansion, nonstationary process, widely linear transformations.**

I. Introduction

In the literature, the use of complex random signals commonly assumes that the signals are proper. A complex zero-mean random signal $s(t)$ is called proper if its second-order properties are completely characterized by the covariance function $\gamma(t_1, t_2) = Es(t_1)s^*(t_2)$. This means its complementary covariance function $c(t_1, t_2) = Es(t_1)s(t_2)$ must be identically zero [1]. Proper signals play a prominent role in communications and signal processing because analytic and equivalent low-pass signals constructed from wide-sense stationary (WSS) real signals are always proper [2]. Therefore, since thermal noise is usually assumed to be WSS, its equivalent low-pass representation is proper. Information-bearing signals, on the other hand, are usually not WSS, and thus possibly improper. Detection and estimation algorithms that assign statistical properties to the signal[1] offer better performance if impropriety of signals is taken into account. In this paper, we solve the problems of detecting and estimating nonstationary improper complex random signals in additive white Gaussian noise. This continues the development of the theory of improper random signals reported in [2]–[6].

An essential tool for the treatment of nonstationary signals is the Karhunen–Loève (K-L) expansion. Because the well-known form of this expansion is only valid in the proper case, we develop the improper version of it in Section II. The K-L expansion is the infinite-dimensional equivalent of the eigenvalue decomposition (EVD) in the finite-dimensional case. The improper version of both the EVD [6, Proposition 1] and the K-L expansion differ from their proper versions in that there are two sets of eigenvalues instead of one and the internal description, in terms of observable coordinates, is improper instead of proper.

Based on the K-L expansion of improper complex signals we are able to solve the problems of estimating and detecting an improper signal in additive white Gaussian noise in Sections IV and V. The basic question is how much can we gain by correctly accounting for the information contained in the complementary covariance, or, equivalently, how much would we lose by treating improper signals in the conventional way. In Section III, we show that, despite the fact that there is a variety of different applications and associated performance measures, there is a surprisingly general answer to this question. In fact, we prove that both for the estimation and the detection problem the performance gain, as measured by mean-squared error (MSE) and deflection [7], respectively, can be as large as 2. In a communications example, we demonstrate how this finding generalizes the result that coherent processing enjoys a 3-dB gain over noncoherent processing. We also show that if the additive noise happens to be improper, the performance advantage can actually be arbitrarily large.

II. Mercer's Theorem and K-L Expansion

The tools required for dealing with improper complex signals have been presented in a unified framework for the vector and WSS cases in [6]. The basic ideas carry over to a nonstationary setting. The appropriate algebra is based on augmented signals $\boldsymbol{\sigma}(t) = [s(t)\ \ s^*(t)]^{\mathsf{T}}$ that carry along their complex conjugate. We shall use the notation $\boldsymbol{\sigma}(t) \Leftrightarrow s(t)$ to indicate that the augmented signal $\boldsymbol{\sigma}(t)$ is an equivalent description of $s(t)$. The covariance matrix of $\boldsymbol{\sigma}(t)$ is called the augmented covariance matrix [8] of $s(t)$ and is given by

$$\boldsymbol{R}(t_1, t_2) = E\boldsymbol{\sigma}(t_1)\boldsymbol{\sigma}^{\mathsf{H}}(t_2) = E\begin{bmatrix} s(t_1) \\ s^*(t_1) \end{bmatrix} [\, s^*(t_2) \quad s(t_2)\,] = \begin{bmatrix} \gamma(t_1, t_2) & c(t_1, t_2) \\ c^*(t_1, t_2) & \gamma^*(t_1, t_2) \end{bmatrix}. \tag{1}$$

The augmented covariance matrix gives a complete second-order description of the signal. It belongs to a matrix algebra

$$\mathcal{W} = \left\{ \begin{bmatrix} h_1(t_1, t_2) & h_2(t_1, t_2) \\ h_2^*(t_1, t_2) & h_1^*(t_1, t_2) \end{bmatrix} \middle| \, h_1, h_2 \in L^2([0, T]^2 \to \mathbb{C}) \right\} \tag{2}$$

which is closed under addition, multiplication, inversion (when inverses exist), and multiplication with a real, but not with a complex scalar. The notation $L^2([0,T]^2 \to \mathbb{C})$ stands for square-integrable

Manuscript received January 29, 2003; revised April 18, 2004. This work was supported by the 1999 National Science Foundation Wireless Initiative under Contract ECS-9979400, the 2001 National Science Foundation ITR Initiative under Contract CCR-0112573, and by the Office of Naval Research under Contract N00014-01-1-1019.

P. J. Schreier is with the School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: peter@peter-schreier.com).

L. L. Scharf is with the Departments of Electrical and Computer Engineering and Statistics, Colorado State University, Ft. Collins, CO 80523-1373 USA (e-mail: scharf@engr.colostate.edu).

C. T. Mullis is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: mullis@schof.colorado.edu).

Communicated by A. Kavčić, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2004.839538

[1]Not all detection strategies assign statistical properties to the signal. For instance, in maximum-likelihood detection, the message is assumed to be unknown, but deterministic. In this case, it is *irrelevant* whether or not the message is improper because the probability density of the observation conditioned on the message is proper if the noise is proper.

functions defined on $[0,T] \times [0,T]$ that take their values in the complex field. The operation

$$r(t_1) = \int_0^T h_1(t_1,t_2)s(t_2)\,dt_2 + \int_0^T h_2(t_1,t_2)s^*(t_2)\,dt_2 \tag{3}$$

is called widely linear (WL) because it depends linearly not only on $s(t)$ but also on $s^*(t)$. Strictly linear (SL) operations are described by $h_2(t_1,t_2) \equiv 0$ and are thus a subset of WL operations.

We now state Mercer's theorem and the K-L expansion for improper complex signals.

Theorem 1: Suppose that $\{s(t),\, 0 \le t \le T\}$ is a zero-mean second-order complex random process characterized by augmented covariance $\boldsymbol{R}(t_1,t_2)$, where both covariance $\gamma(t_1,t_2)$ and complementary covariance $c(t_1,t_2)$ are continuous on $[0,T]^2$. Then $\boldsymbol{R}(t_1,t_2)$ can be expanded in the uniformly and absolutely convergent series

$$\boldsymbol{R}(t_1,t_2) = \sum_n \boldsymbol{\Phi}_n(t_1)\boldsymbol{\Lambda}_n\boldsymbol{\Phi}_n^{\mathsf{H}}(t_2). \tag{4}$$

The matrix $\boldsymbol{\Lambda}_n$ is real and it involves two nonnegative eigenvalues λ_n^r and λ_n^i

$$\boldsymbol{\Lambda}_n = \frac{1}{2}\begin{bmatrix} \lambda_n^r + \lambda_n^i & \lambda_n^r - \lambda_n^i \\ \lambda_n^r - \lambda_n^i & \lambda_n^r + \lambda_n^i \end{bmatrix}. \tag{5}$$

The matrix

$$\boldsymbol{\Phi}_n(t) = \begin{bmatrix} \phi_n(t) & \psi_n(t) \\ \psi_n^*(t) & \phi_n^*(t) \end{bmatrix} \in \mathcal{W} \tag{6}$$

satisfies the orthogonality condition

$$\int_0^T \boldsymbol{\Phi}_n(t)\boldsymbol{\Phi}_m^{\mathsf{H}}(t)\,dt = \boldsymbol{I}\,\delta_{nm}. \tag{7}$$

The matrices $\boldsymbol{\Lambda}_n$ and $\boldsymbol{\Phi}_n(t)$ are found as the solutions to the equation

$$\boldsymbol{\Phi}_n(t_1)\boldsymbol{\Lambda}_n = \int_0^T \boldsymbol{R}(t_1,t_2)\boldsymbol{\Phi}_n(t_2)\,dt_2. \tag{8}$$

Then $\boldsymbol{\sigma}(t) \Leftrightarrow s(t)$ can be represented by the mean-square convergent series

$$\boldsymbol{\sigma}(t) = \sum_n \boldsymbol{\Phi}_n(t)\boldsymbol{\sigma}_n \Leftrightarrow s(t) = \sum_n \phi_n(t)s_n + \psi_n(t)s_n^* \tag{9}$$

where the complex K-L coefficients are

$$\begin{aligned} \boldsymbol{\sigma}_n &= \int_0^T \boldsymbol{\Phi}_n^{\mathsf{H}}(t)\boldsymbol{\sigma}(t)dt \\ \Leftrightarrow \quad s_n &= \int_0^T (\phi_n^*(t)s(t) + \psi_n(t)s^*(t))\,dt. \end{aligned} \tag{10}$$

These K-L coefficients s_n are improper, with the following covariance and complementary covariance:

$$E(s_n s_m^*) = \frac{1}{2}(\lambda_n^r + \lambda_n^i)\delta_{nm} \tag{11}$$

$$E(s_n s_m) = \frac{1}{2}(\lambda_n^r - \lambda_n^i)\delta_{nm}. \tag{12}$$

Proof: Let $\mathbb{C}_*^2$ be the image of $\mathbb{R}^2$ under the unitary map

$$\boldsymbol{T} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix}. \tag{13}$$

The space of augmented signals $\boldsymbol{\sigma}(t)$ consists of square-integrable functions $[0,T] \to \mathbb{C}_*^2$ and is denoted by $L^2([0,T] \to \mathbb{C}_*^2)$. This space is not linear (rather it is WL), but it is isomorphic to the space of square-integrable maps $[0,T] \to \mathbb{R}^2$. This isomorphism enables us to write down the K-L expansion for an augmented signal, using the results of [16] for vector random processes. Let the assumptions be as in the statement of the theorem. Then the augmented covariance $\boldsymbol{R}(t_1,t_2)$ can be expanded in the series

$$\boldsymbol{R}(t_1,t_2) = \sum_n \lambda_n \boldsymbol{\varphi}_n(t_1)\boldsymbol{\varphi}_n^{\mathsf{H}}(t_2) \tag{14}$$

where $\{\lambda_n\}$ are the nonnegative scalar eigenvalues and the

$$\{\boldsymbol{\varphi}_n(t) = [f_n(t), f_n^*(t)]^{\mathsf{T}}\}$$

are the corresponding orthonormal eigenfunctions. Each $\boldsymbol{\varphi}_n(t)$ is $L^2([0,T] \to \mathbb{C}_*^2)$. Eigenvalues and eigenfunctions are obtained as solutions to the integral equation

$$\lambda_n\boldsymbol{\varphi}_n(t_1) = \int_0^T \boldsymbol{R}(t_1,t_2)\boldsymbol{\varphi}_n(t_2)dt_2, \qquad 0 \le t_1 \le T \tag{15}$$

where the eigenfunctions $\boldsymbol{\varphi}_n(t)$ form a complete orthonormal set for $L^2([0,T] \to \mathbb{C}_*^2)$

$$\int_0^T \boldsymbol{\varphi}_n^{\mathsf{H}}(t)\boldsymbol{\varphi}_m(t)dt = 2\,\mathrm{Re}\int_0^T f_n^*(t)f_m(t)\,dt = \delta_{nm}. \tag{16}$$

Then $\boldsymbol{\sigma}(t) \Leftrightarrow s(t)$ can be represented by the mean-square convergent series

$$\boldsymbol{\sigma}(t) = \sum_n x_n\boldsymbol{\varphi}_n(t) \Leftrightarrow s(t) = \sum_n x_n f_n(t) \tag{17}$$

where the x_n are the K-L coefficients

$$x_n = \int_0^T \boldsymbol{\varphi}_n^{\mathsf{H}}(t)\boldsymbol{\sigma}(t)dt = 2\,\mathrm{Re}\int_0^T f_n^*(t)s(t)\,dt. \tag{18}$$

The surprising result here is that the K-L coefficients x_n are *real* scalars with correlation

$$E\,x_n x_m = \lambda_n\delta_{nm}. \tag{19}$$

The reason for this lies in (16). This equation shows that the functions $f_n(t)$ do not have to be orthogonal in $L^2([0,T] \to \mathbb{C})$ to ensure that the eigenfunctions $\boldsymbol{\varphi}_n(t)$ be orthogonal in $L^2([0,T] \to \mathbb{C}_*^2)$. In fact, it is clear that in general there are more orthogonal augmented functions $\boldsymbol{\varphi}_n(t)$ in $L^2([0,T] \to \mathbb{C}_*^2)$ than there are orthogonal functions $f_n(t)$ in $L^2([0,T] \to \mathbb{C})$. In other words, we are able to reduce the dimension of the internal description (real rather than complex K-L coefficients) because $L^2([0,T] \to \mathbb{C}_*^2)$ allows more orthogonal eigenfunctions than $L^2([0,T] \to \mathbb{C})$. This increase in the number of eigenfunctions is not clearly visible since in both cases we have infinitely many.

From (14) it is not clear how the improper version of Mercer's theorem is connected to its proper version. To make this connection apparent, we rewrite (14) as

$$\begin{aligned} &\boldsymbol{R}(t_1,t_2) \\ &= \sum_n \left([\boldsymbol{\varphi}_{2n}(t_1), \boldsymbol{\varphi}_{2n+1}(t_1)]\boldsymbol{T}^{\mathsf{H}}\right) \\ &\quad \times \left(\boldsymbol{T}\begin{bmatrix} \lambda_{2n} & 0 \\ 0 & \lambda_{2n+1} \end{bmatrix}\boldsymbol{T}^{\mathsf{H}}\right)\left(\boldsymbol{T}\begin{bmatrix} \boldsymbol{\varphi}_{2n}^{\mathsf{H}}(t_2) \\ \boldsymbol{\varphi}_{2n+1}^{\mathsf{H}}(t_2) \end{bmatrix}\right) \\ &= \sum_n \begin{bmatrix} \phi_n(t_1) & \psi_n(t_1) \\ \psi_n^*(t_1) & \phi_n^*(t_1) \end{bmatrix} \\ &\quad \times \begin{bmatrix} \frac{1}{2}(\lambda_{2n}+\lambda_{2n+1}) & \frac{1}{2}(\lambda_{2n}-\lambda_{2n+1}) \\ \frac{1}{2}(\lambda_{2n}-\lambda_{2n+1}) & \frac{1}{2}(\lambda_{2n}+\lambda_{2n+1}) \end{bmatrix}\begin{bmatrix} \phi_n^*(t_2) & \psi_n(t_2) \\ \psi_n^*(t_2) & \phi_n(t_2) \end{bmatrix} \\ &= \sum_n \boldsymbol{\Phi}_n(t_1)\boldsymbol{\Lambda}_n\boldsymbol{\Phi}_n^{\mathsf{H}}(t_2) \end{aligned} \tag{20}$$

where $\phi_n(t) = f_{2n}(t) - jf_{2n+1}(t)$ and $\psi_n(t) = f_{2n}(t) + jf_{2n+1}(t)$. Thus, the internal representation is now given by *complex* K-L coefficients $s_n = \frac{1}{\sqrt{2}}(x_{2n} + jx_{2n+1})$

$$\boldsymbol{\sigma}_n = \begin{bmatrix} s_n \\ s_n^* \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} x_{2n} \\ x_{2n+1} \end{bmatrix} = \int_0^T \boldsymbol{\Phi}_n^{\mathsf{H}}(t)\boldsymbol{\sigma}(t)\,dt. \tag{21}$$

For these coefficients, we find because of (19)

$$\begin{aligned} E(s_n s_m) &= \frac{1}{2}[E(x_{2n}x_{2m}) - E(x_{2n+1}x_{2m+1}) \\ &\quad + j(E(x_{2n}x_{2m+1}) + E(x_{2n+1}x_{2m}))] \\ &= \frac{1}{2}(\lambda_{2n} - \lambda_{2n+1})\delta_{nm} \end{aligned} \tag{22}$$

and, similarly

$$E(s_n s_m^*) = \frac{1}{2}(\lambda_{2n} + \lambda_{2n+1})\delta_{nm}.$$

If we now agree on the definitions $\lambda_n^r \triangleq \lambda_{2n}$, $\lambda_n^i \triangleq \lambda_{2n+1}$, the proof is complete. □

This theorem is the continuous-time equivalent of the finite-dimensional eigenvalue decomposition for the improper case presented as Proposition 1 in [6]. The main differences between the proper and improper version of this theorem are that there are two sets of eigenvalues $\{\lambda_n^r\}$, $\{\lambda_n^i\}$, and the K-L coefficients s_n are improper. In the proper case, $c(t_1, t_2) \equiv 0$, the two eigenvalues λ_n^r and λ_n^i are equal, the K-L coefficients s_n become proper, and the functions $\psi_n(t) \equiv 0\ \forall n$. Then the K-L expansion of Theorem 1 simplifies to the known formulas.

Example: To gain more insight into the role of the two sets of eigenvalues, consider the following communications example. Suppose we want to detect a *real* waveform $x(t)$ that is transmitted over a channel that rotates it by some random phase ϕ and adds complex white Gaussian noise $n(t)$. The observations are then given by

$$r(t) = x(t)e^{j\phi} + n(t) \tag{23}$$

where we shall assume mutual independence of $x(t)$, $n(t)$, and ϕ. Furthermore, denote the rotated signal by $s(t) = x(t)e^{j\phi}$. Its covariance is given by $\gamma(t_1, t_2) = Ex(t_1)x(t_2)$ and its complementary covariance is $c(t_1, t_2) = Es(t_1)s(t_2) = Ex(t_1)x(t_2) \cdot Ee^{j2\phi}$.

There are, of course, two important special cases. If the phase ϕ is uniformly distributed, then $c(t_1, t_2) \equiv 0$ and detection will be *noncoherent*. The eigenvalues of the augmented covariance of $s(t)$ satisfy $\lambda_n^r = \lambda_n^i = \mu_n$. On the other hand, if ϕ is known, then $c(t_1, t_2) \equiv e^{j2\phi}\gamma(t_1, t_2)$ and detection will be *coherent*. If we order the eigenvalues of the augmented covariance of $s(t)$ appropriately, we have $\lambda_n^r = 2\mu_n$ and $\lambda_n^i = 0$. Therefore, the coherent case is the most improper case under the power constraint $\lambda_n^r + \lambda_n^i = 2\mu_n$. These comments are clarified by noting that

$$\frac{1}{2}\lambda_n^r = E\{s(t)\mathrm{Re}\, s_n\}, \qquad \frac{j}{2}\lambda_n^i = E\{s(t)\mathrm{Im}\, s_n\}. \tag{24}$$

Thus, λ_n^r measures the correlation between the real part of the observable coordinate s_n and the continuous-time signal, and λ_n^i does so for the imaginary part. In the noncoherent version of (23), these two correlations are equal, suggesting that the information is carried equally in the real and imaginary parts of s_n. In the coherent version, $\lambda_n^r = 2\mu_n$, $\lambda_n^i = 0$ shows that the information is carried exclusively in the real part of s_n, making $\mathrm{Re}\, s_n$ a *sufficient statistic* for the decision on $s(t)$. Therefore, in the coherent problem, WL processing amounts to only considering the real part of the internal description. The more interesting applications of WL filtering, however, lie either in between the coherent and noncoherent case, characterized by a nonuniform phase distribution, or in adaptive realizations of coherent algorithms.

III. Performance Comparisons Between WL and SL Processing

It is of particular interest to compare the performance achieved by WL processing with that of SL processing. There is obviously a variety of different applications and, thus, a variety of different performance measures, and at this point it is not clear whether we can hope for a unifying approach. Many of these performance measures, however, are of the class $M\left(\{\lambda_n^r, \lambda_n^i\}, \boldsymbol{p}\right)$; i.e., they are functions of the eigenvalues $\{\lambda_n^r, \lambda_n^i\}$ of the augmented covariance of $s(t)$ and $\boldsymbol{p}$, a vector of parameters. Typically, $\boldsymbol{p}$ contains at least the noise level N_0. For this particular class of performance measures, there is a very general result comparing WL with SL processing, which we now derive.

To make the following discussion as simple as possible, we shall work with an N-dimensional signal $\boldsymbol{s}$, which is obtained from $s(t)$ by uniformly sampling it in $[0, T]$. The sampled signal $\boldsymbol{s}$ has covariance matrix $\boldsymbol{\Gamma} = E\boldsymbol{s}\boldsymbol{s}^{\mathsf{H}}$ and complementary covariance matrix $\boldsymbol{C} = E\boldsymbol{s}\boldsymbol{s}^{\mathsf{T}}$. The K-L expansion of $s(t)$ is then approximated by the eigenvalue decomposition of $\boldsymbol{s}$ [6, Proposition 1]. Since the K-L expansion produces a *countably* infinite number of eigenvalues, the approximation can be made as good as necessary by making the sampling grid finer and the corresponding augmented covariance matrix $\boldsymbol{R}$ larger.

Let

$$\boldsymbol{R}_{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{C} \\ \boldsymbol{C}^* & \boldsymbol{\Gamma}^* \end{bmatrix} \tag{25}$$

denote the $2N \times 2N$ augmented covariance matrix of $\boldsymbol{s}$ with fixed $\boldsymbol{\Gamma}$ and free $\boldsymbol{C}$. The set of all $\boldsymbol{R}_{\boldsymbol{C}}$ will be called $\mathcal{R}_{\boldsymbol{C}}$. Since every matrix $\boldsymbol{R}_{\boldsymbol{C}}$ must be positive semidefinite, the set $\mathcal{R}_{\boldsymbol{C}}$ is convex and compact. Assuming that greater $M(\cdot)$ means better performance, the maximum performance gain of WL over SL processing can be expressed as

$$\max M_{\mathsf{WL}}/M_{\mathsf{SL}} = \max_{\boldsymbol{p}} \left[\max_{\boldsymbol{R}_{\boldsymbol{C}} \in \mathcal{R}_{\boldsymbol{C}}} \frac{M(\boldsymbol{ev}(\boldsymbol{R}_{\boldsymbol{C}}), \boldsymbol{p})}{M(\boldsymbol{ev}(\boldsymbol{R}_{\boldsymbol{0}}), \boldsymbol{p})}\right] \tag{26}$$

where $\boldsymbol{ev}(\boldsymbol{R}_{\boldsymbol{C}})$ denotes the vector of eigenvalues of $\boldsymbol{R}_{\boldsymbol{C}}$, and $\boldsymbol{ev}(\boldsymbol{R}_{\boldsymbol{0}})$ the vector of eigenvalues of $\boldsymbol{R}_{\boldsymbol{C}}$ with $\boldsymbol{C} = \boldsymbol{0}$. Also, from (25) it is clear that $\boldsymbol{ev}(\boldsymbol{R}_{\boldsymbol{0}}) = [\boldsymbol{\mu}, \boldsymbol{\mu}]$, where $\boldsymbol{\mu} = \boldsymbol{ev}(\boldsymbol{\Gamma})$. The question is what choice of $\boldsymbol{C}$ maximizes the performance criterion $M(\cdot)$. To answer this question, we need some definitions and intermediate results.

Definition 1 (Majorization) [9, p. 7]: A real $N \times 1$ vector $\boldsymbol{x}$ is said to be *majorized* by a real $N \times 1$ vector $\boldsymbol{y}$, written as $\boldsymbol{x} \prec \boldsymbol{y}$, if

$$\sum_{k=1}^{n} x_{[k]} \leq \sum_{k=1}^{n} y_{[k]}, \qquad n = 1, \ldots, N-1 \tag{27}$$

$$\sum_{k=1}^{N} x_{[k]} = \sum_{k=1}^{N} y_{[k]}. \tag{28}$$

where $[\cdot]$ is a permutation operator such that $x_{[1]} \geq \cdots \geq x_{[N]}$.

Intuitively, if $\boldsymbol{x} \prec \boldsymbol{y}$, then the components of $\boldsymbol{x}$ are "less spread out" or "more equal" than the components of $\boldsymbol{y}$. Note that majorization is sometimes also defined with a permutation operator that arranges the components of $\boldsymbol{x}$ in *increasing* order [10, Definition 4.3.24], $x_{[1]} \leq \cdots \leq x_{[N]}$.

Somewhat loosely stated, we will now show the following. For a given covariance $\boldsymbol{\Gamma}$, the signal whose augmented covariance has least

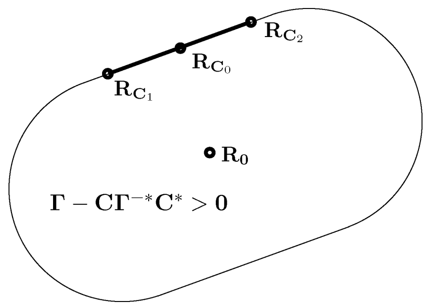

Fig. 1. The convex and compact set $\mathcal{R}_{\boldsymbol{C}}$. The interior of this set is described by $\boldsymbol{P_C} = \boldsymbol{\Gamma} - \boldsymbol{C}\boldsymbol{\Gamma}^{-*}\boldsymbol{C}^* > \mathbf{0}$. Boundary points are rank-deficient points. However, boundary points with $\boldsymbol{P_C} \geq \mathbf{0}$ but $\boldsymbol{P_C} \neq \mathbf{0}$ are not extreme points. For instance, the boundary point $\boldsymbol{R}_{\boldsymbol{C}_0}$ shown in this figure is a convex combination of the points $\boldsymbol{R}_{\boldsymbol{C}_1}$ and $\boldsymbol{R}_{\boldsymbol{C}_2}$.

eigenvalue spread must be proper, i.e., have $\boldsymbol{C} = \mathbf{0}$. The more improper a signal becomes the more spread out become the eigenvalues of its augmented covariance. Finally, the most improper signal, again for a given covariance $\boldsymbol{\Gamma}$, corresponds to an extreme point $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$ in the convex set $\mathcal{R}_{\boldsymbol{C}}$. The geometry of $\mathcal{R}_{\boldsymbol{C}}$ is depicted in Fig. 1.

Lemma 1: If $\boldsymbol{\Gamma}$ is invertible, a point $\boldsymbol{R}_{\hat{\boldsymbol{C}}} \in \mathcal{R}_{\boldsymbol{C}}$ is an extreme point of $\mathcal{R}_{\boldsymbol{C}}$ if and only if the Schur complement of $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$ vanishes: $\boldsymbol{P}_{\hat{\boldsymbol{C}}} = \boldsymbol{\Gamma} - \hat{\boldsymbol{C}}\boldsymbol{\Gamma}^{-*}\hat{\boldsymbol{C}}^* = \mathbf{0}$. In other words, $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$ is an extreme point of $\mathcal{R}_{\boldsymbol{C}}$ if and only if it has minimum rank for given $\boldsymbol{\Gamma}$.

Proof: An extreme point $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$ of the closed convex set $\mathcal{R}_{\boldsymbol{C}}$ is a point in $\mathcal{R}_{\boldsymbol{C}}$ that may be written as a convex combination of points from $\mathcal{R}_{\boldsymbol{C}}$ in only a trivial way [10, Appendix B]; that is, $\boldsymbol{R}_{\hat{\boldsymbol{C}}} = \alpha\boldsymbol{R}_{\boldsymbol{C}_1} + \bar{\alpha}\boldsymbol{R}_{\boldsymbol{C}_2}$, $0 < \alpha < 1$, $\bar{\alpha} = 1 - \alpha$, and $\boldsymbol{R}_{\boldsymbol{C}_1}, \boldsymbol{R}_{\boldsymbol{C}_2} \in \mathcal{R}_{\boldsymbol{C}}$, implies that $\boldsymbol{R}_{\boldsymbol{C}_1} = \boldsymbol{R}_{\boldsymbol{C}_2} = \boldsymbol{R}_{\hat{\boldsymbol{C}}}$.

We first show that a point $\boldsymbol{R}_{\boldsymbol{C}_0} \in \mathcal{R}_{\boldsymbol{C}}$ with nonzero Schur complement $\boldsymbol{P}_{\boldsymbol{C}_0}$ cannot be an extreme point. Note that we can assume without loss of generality that $\boldsymbol{\Gamma} = \boldsymbol{I}$. Then every point $\boldsymbol{R}_{\boldsymbol{C}} \in \mathcal{R}_{\boldsymbol{C}}$ must satisfy[2] $\boldsymbol{C}\boldsymbol{C}^{\mathsf{H}} \leq \boldsymbol{I}$, which means that all singular values of $\boldsymbol{C}$, denoted by $\sigma_i, i = 1, \dots, N$, must satisfy $\sigma_i \leq 1$. The matrix $\boldsymbol{C}$ is symmetric and therefore has two singular value decompositions (SVDs)

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{V}^*\boldsymbol{S}\boldsymbol{U}^{\mathsf{T}} = \boldsymbol{C}^{\mathsf{T}}. \tag{29}$$

Since

$$\begin{aligned} \boldsymbol{C}\boldsymbol{C}^{\mathsf{H}} &= \boldsymbol{U}\boldsymbol{S}^2\boldsymbol{U}^{\mathsf{H}} = \boldsymbol{V}^*\boldsymbol{S}^2\boldsymbol{V}^{\mathsf{T}} \\ &\Leftrightarrow \boldsymbol{S}^2\boldsymbol{V}^{\mathsf{T}}\boldsymbol{U} = \boldsymbol{V}^{\mathsf{T}}\boldsymbol{U}\boldsymbol{S}^2 \\ &\Leftrightarrow \boldsymbol{S}\boldsymbol{V}^{\mathsf{T}}\boldsymbol{U} = \boldsymbol{V}^{\mathsf{T}}\boldsymbol{U}\boldsymbol{S} \end{aligned} \tag{30}$$

the unitary matrix $\boldsymbol{T} = \boldsymbol{V}^{\mathsf{T}}\boldsymbol{U}$ commutes with $\boldsymbol{S}$. Therefore, whenever $\sigma_i \neq \sigma_j, T_{ij} = 0$. From this development, we see that every $\boldsymbol{C}$ can be expressed as

$$\boldsymbol{C} = \boldsymbol{V}^*\boldsymbol{D}\boldsymbol{V}^{\mathsf{H}} \tag{31}$$

where $\boldsymbol{D} = \boldsymbol{T}\boldsymbol{S}$ is block-diagonal with diagonal blocks of the form $\boldsymbol{D}_i = \sigma_i\boldsymbol{W}_i$, where σ_i is a singular value of $\boldsymbol{C}$ and $\boldsymbol{W}_i$ is unitary and symmetric.

If $\boldsymbol{P}_{\boldsymbol{C}_0} \neq \mathbf{0}$, there is at least one singular value, say σ_i, with $0 \leq \sigma_i < 1$. Then set $\alpha = (1 - \sigma_i)/2$ so that $\sigma_i\boldsymbol{W}_i = \alpha(-\boldsymbol{W}_i) + \bar{\alpha}\boldsymbol{W}_i$. Thus, we construct $\boldsymbol{C}_1 = \boldsymbol{V}^*\boldsymbol{D}'\boldsymbol{V}^{\mathsf{H}}$ and $\boldsymbol{C}_2 = \boldsymbol{V}^*\boldsymbol{D}''\boldsymbol{V}^{\mathsf{H}}$, where $\boldsymbol{D}'$ and $\boldsymbol{D}''$ agree with $\boldsymbol{D}$ except on the block $\boldsymbol{D}_i$ corresponding to singular value σ_i. We replace σ_i with -1 to get $\boldsymbol{C}_1$ and with 1 to get $\boldsymbol{C}_2$. This construction guarantees that $\boldsymbol{R}_{\boldsymbol{C}_1}, \boldsymbol{R}_{\boldsymbol{C}_2} \in \mathcal{R}_{\boldsymbol{C}}$. It shows that $\boldsymbol{R}_{\boldsymbol{C}_0}$ can be written as a nontrivial convex combination of $\boldsymbol{R}_{\boldsymbol{C}_1}$ and $\boldsymbol{R}_{\boldsymbol{C}_2}$: $\boldsymbol{R}_{\boldsymbol{C}_0} = \alpha\boldsymbol{R}_{\boldsymbol{C}_1} + \bar{\alpha}\boldsymbol{R}_{\boldsymbol{C}_2}$. This is depicted in Fig. 1.

Assume now that $\boldsymbol{P}_{\hat{\boldsymbol{C}}} = \mathbf{0}$ for some point $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$. The Schur complement is a matrix-concave function, which means that

$$\boldsymbol{P}_{\hat{\boldsymbol{C}}} \geq \alpha\boldsymbol{P}_{\boldsymbol{C}_1} + \bar{\alpha}\boldsymbol{P}_{\boldsymbol{C}_2} \tag{32}$$

for $\hat{\boldsymbol{C}} = \alpha\boldsymbol{C}_1 + \bar{\alpha}\boldsymbol{C}_2$ with $\boldsymbol{R}_{\boldsymbol{C}_1}, \boldsymbol{R}_{\boldsymbol{C}_2} \in \mathcal{R}_{\boldsymbol{C}}$. Equality in (32) holds if and only if $\boldsymbol{C}_1 = \boldsymbol{C}_2$. The matrix-concavity is an immediate consequence of $\alpha\bar{\alpha}(\boldsymbol{C}_1 - \boldsymbol{C}_2)\boldsymbol{\Gamma}^{-*}(\boldsymbol{C}_1 - \boldsymbol{C}_2)^{\mathsf{H}} \geq \mathbf{0}$, which is an equality if and only if $\boldsymbol{C}_1 = \boldsymbol{C}_2$. Since $\boldsymbol{P}_{\boldsymbol{C}_1} \geq \mathbf{0}$ and $\boldsymbol{P}_{\boldsymbol{C}_2} \geq \mathbf{0}$, a vanishing Schur complement $\boldsymbol{P}_{\hat{\boldsymbol{C}}} = \mathbf{0}$ requires $\boldsymbol{P}_{\boldsymbol{C}_1} = \boldsymbol{P}_{\boldsymbol{C}_2} = \mathbf{0}$ and $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \hat{\boldsymbol{C}}$. Thus, $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$ is indeed an extreme point of $\mathcal{R}_{\boldsymbol{C}}$. □

It is trivial to show that the eigenvalues of all extreme points are $[2\boldsymbol{\mu}, \mathbf{0}_N]$, $\mathbf{0}_N$ being the $N \times 1$ null vector. We are now in a position to precisely state one of our main results.

Lemma 2: Within the set $\mathcal{R}_{\boldsymbol{C}}$, we have the following partial ordering:

$$\boldsymbol{ev}(\boldsymbol{R_0}) = [\boldsymbol{\mu}, \boldsymbol{\mu}] \prec \boldsymbol{ev}(\boldsymbol{R_C}) \prec [2\boldsymbol{\mu}, \mathbf{0}_N] = \boldsymbol{ev}(\boldsymbol{R}_{\hat{\boldsymbol{C}}}) \tag{33}$$

where $\boldsymbol{R}_{\hat{\boldsymbol{C}}}$ is any extreme point of $\mathcal{R}_{\boldsymbol{C}}$.

Proof: The left inequality has been proved in [6, Proposition 4]. The right inequality is a consequence of a theorem by Thompson and Therianos [11], also discussed in [9, Sec. 9.C.4]. Applied to the matrix $\boldsymbol{R_C}$, the theorem says that

$$\sum_{k=1}^{n} \mathrm{ev}_{[k]}(\boldsymbol{R_C}) + \mathrm{ev}_{[2N-n+k]}(\boldsymbol{R_C}) \leq \sum_{k=1}^{n} 2\mu_{[k]}, \qquad n = 1, \dots, N. \tag{34}$$

To show majorization, we have to demonstrate that

$$\sum_{k=1}^{n} \mathrm{ev}_{[k]}(\boldsymbol{R_C}) \leq \sum_{k=1}^{n} 2\mu_{[k]}, \qquad n = 1, \dots, N \tag{35}$$

$$\sum_{k=1}^{n} \mathrm{ev}_{[k]}(\boldsymbol{R_C}) \leq \sum_{k=1}^{N} 2\mu_{[k]}, \qquad n = N+1, \dots, 2N. \tag{36}$$

The first inequality is an immediate consequence of (34) because all eigenvalues are nonnegative. The second inequality is clear from the trace constraint $\operatorname{tr} \boldsymbol{R_C} = 2 \operatorname{tr} \boldsymbol{\Gamma}$. □

Referring to the communications example (23), we identify noncoherent detection with the matrix $\boldsymbol{R_0}$ and coherent detection with $\boldsymbol{R_\Gamma}$ (with real $\boldsymbol{\Gamma}$), which is an extreme point of $\mathcal{R}_{\boldsymbol{C}}$. To now be able to make a performance comparison based on Lemma 2, the performance measure $M(\cdot)$ must preserve the partial ordering in (33). Functions that preserve a partial ordering by majorization are called *Schur-convex* or *Schur-concave*:

Definition 2 (Schur-Convex and Schur-Concave Function) [9, Definition 3.A.1]: A real-valued function g defined on a set $D \subset \mathbb{R}^N$ is said to be *Schur-convex on* D if $\boldsymbol{x} \prec \boldsymbol{y}$ on D implies that $g(\boldsymbol{x}) \leq g(\boldsymbol{y})$. Similarly, a function is called *Schur-concave on* D if $\boldsymbol{x} \prec \boldsymbol{y}$ on D implies that $g(\boldsymbol{x}) \geq g(\boldsymbol{y})$.

With this definition, we have now proved the following theorem.

Theorem 2: Let $\boldsymbol{R_C}$ be the augmented covariance matrix of a signal to be treated by some system S and let $M(\boldsymbol{ev}(\boldsymbol{R_C}), \boldsymbol{p})$ measure the performance of S. If $M(\cdot)$ is a Schur-convex function on the set of eigenvalues of $\boldsymbol{R_C}$, with $\boldsymbol{p}$ a set of parameters, then the maximum perfor-

[2]By $\boldsymbol{A} > \boldsymbol{B}$ we mean that $\boldsymbol{A} - \boldsymbol{B}$ is a positive-definite matrix. Analogously, $\boldsymbol{A} \geq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

mance gain of widely linear processing over strictly linear processing is, under the criterion $M(\cdot)$

$$\max \frac{M_{\mathsf{WL}}}{M_{\mathsf{SL}}} = \max_{\boldsymbol{p}} \frac{M([2\boldsymbol{\mu}, \mathbf{0}_N], \boldsymbol{p})}{M([\boldsymbol{\mu}, \boldsymbol{\mu}], \boldsymbol{p})}. \tag{37}$$

Let the assumptions be as above, but $M(\boldsymbol{ev}(\boldsymbol{R_C}), \boldsymbol{p})$ be Schur-concave. Then the maximum performance gain is

$$\max \frac{M_{\mathsf{SL}}}{M_{\mathsf{WL}}} = \max_{\boldsymbol{p}} \frac{M([\boldsymbol{\mu}, \boldsymbol{\mu}], \boldsymbol{p})}{M([2\boldsymbol{\mu}, \mathbf{0}_N], \boldsymbol{p})}. \tag{38}$$

There are a number of results that enable us to check whether a function is Schur-convex or Schur-concave. We now simply state some of the results we will use later on; for a detailed discussion of Schur-convex functions we refer the reader to [9].

Fact 1 [9, Sec. 3.A.4]: Let $I \in \mathbb{R}$ be an open interval and $f : I^N \to \mathbb{R}$ be continuously differentiable. Necessary and sufficient conditions for f to be Schur-convex on I^N are: f is symmetric in its arguments on I^N, and for all $\boldsymbol{x} \in I^N$

$$(x_1 - x_2)\left[\frac{\partial}{\partial x_1} f(\boldsymbol{x}) - \frac{\partial}{\partial x_2} f(\boldsymbol{x})\right] \geq 0. \tag{39}$$

For Schur-concave functions this inequality must be reversed.

Fact 2 [9, Sec. 3.C.1]: If $I \in \mathbb{R}$ is an interval and $g : I \to \mathbb{R}$ is convex (concave), then

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} g(x_n) \tag{40}$$

is Schur-convex (Schur-concave) on I^N.

We will see that interesting performance measures such as MSE or deflection are in fact Schur-concave and Schur-convex functions of the eigenvalues of $\boldsymbol{R_C}$.

IV. Estimation

We are interested in estimating a nonstationary improper complex zero-mean random signal $s(t)$, characterized by its augmented covariance $\boldsymbol{R}(t_1, t_2)$, in complex white (i.e., proper) Gaussian noise $n(t)$ with power spectral density N_0. The observations are given by

$$r(t) = s(t) + n(t), \qquad 0 \leq t \leq T \tag{41}$$

and the noise will be assumed to be uncorrelated with the signal.

A. Solution of the Estimation Problem

We are looking for a WL estimator $\hat{\boldsymbol{\sigma}}(t) \Leftrightarrow \hat{s}(t)$ of the form

$$\hat{\boldsymbol{\sigma}}(t) = \int_0^T \boldsymbol{H}(t,v)\boldsymbol{\rho}(v)dv \tag{42}$$

where $\boldsymbol{H}(t,v) \in \mathcal{W}$, $\boldsymbol{\rho}(v) = [r(v), r^*(v)]^{\mathsf{T}} \Leftrightarrow r(v)$, and thus,

$$\hat{s}(t) = \int_0^T h_1(t,v)r(v)dv + \int_0^T h_2(t,v)r^*(v)dv. \tag{43}$$

To make this estimator a minimum MSE (MMSE) estimator, we require it to satisfy the orthogonality condition

$$\boldsymbol{\sigma}(t) - \hat{\boldsymbol{\sigma}}(t) \perp \boldsymbol{\rho}(u) \tag{44}$$

which translates to

$$\boldsymbol{R}(t,u) = \int_0^T \boldsymbol{H}(t,v) E\boldsymbol{\rho}(v)\boldsymbol{\rho}^{\mathsf{H}}(u)\,dv. \tag{45}$$

Using the K-L expansion of (4) in this equation we obtain

$$\sum_n \boldsymbol{\Phi}_n(t)\boldsymbol{\Lambda}_n\boldsymbol{\Phi}_n^{\mathsf{H}}(u) = \int_0^T \boldsymbol{H}(t,v)\sum_n \boldsymbol{\Phi}_n(v)(\boldsymbol{\Lambda}_n + N_0\boldsymbol{I})\boldsymbol{\Phi}_n^{\mathsf{H}}(u)\,dv. \tag{46}$$

We now attempt a solution of the form

$$\boldsymbol{H}(t,v) = \sum_n \boldsymbol{\Phi}_n(t)\boldsymbol{H}_n\boldsymbol{\Phi}_n^{\mathsf{H}}(v) \tag{47}$$

$$\boldsymbol{H}_n = \begin{bmatrix} h_{n,1} & h_{n,2} \\ h_{n,2}^* & h_{n,1}^* \end{bmatrix} \in \mathcal{W}. \tag{48}$$

Inserting (47) into (46) we get

$$\sum_n \boldsymbol{\Phi}_n(t)\boldsymbol{\Lambda}_n\boldsymbol{\Phi}_n^{\mathsf{H}}(u) = \sum_n \boldsymbol{\Phi}_n(t)\boldsymbol{H}_n(\boldsymbol{\Lambda}_n + N_0\boldsymbol{I})\boldsymbol{\Phi}_n^{\mathsf{H}}(u) \tag{49}$$

which means

$$\boldsymbol{H}_n = \boldsymbol{\Lambda}_n(\boldsymbol{\Lambda}_n + N_0\boldsymbol{I})^{-1}. \tag{50}$$

Thus, the terms of $\boldsymbol{H}_n$ are

$$h_{n,1} = \frac{\lambda_n^r\lambda_n^i + \frac{N_0}{2}(\lambda_n^r + \lambda_n^i)}{\lambda_n^r\lambda_n^i + N_0(\lambda_n^r + \lambda_n^i) + N_0^2} \tag{51}$$

$$h_{n,2} = \frac{\frac{N_0}{2}(\lambda_n^r - \lambda_n^i)}{\lambda_n^r\lambda_n^i + N_0(\lambda_n^r + \lambda_n^i) + N_0^2}. \tag{52}$$

If we denote the resolution of the observed signal onto the K-L basis functions by

$$\boldsymbol{\rho}_n = \int_0^T \boldsymbol{\Phi}_n^{\mathsf{H}}(t)\boldsymbol{\rho}(t)\,dt \Leftrightarrow r_n = \int_0^T \phi_n^*(t)r(t) + \psi_n(t)r^*(t)\,dt \tag{53}$$

the WL estimator in (42) is

$$\hat{\boldsymbol{\sigma}}(t) = \sum_n \boldsymbol{\Phi}_n(t)\boldsymbol{H}_n\boldsymbol{\rho}_n \Leftrightarrow \hat{s}(t) = \sum_n \hat{s}_n\phi_n(t) + \hat{s}_n^*\psi_n(t) \tag{54}$$

$$\hat{s}_n = h_{n,1}r_n + h_{n,2}r_n^*. \tag{55}$$

In the proper case, $\lambda_n^r = \lambda_n^i = \mu_n$, $\psi_n(t) \equiv 0\ \forall n$, and we have $h_{n,1} = \frac{\mu_n}{\mu_n + N_0}$ and $h_{n,2} = 0$. Thus, the solution simplifies to the known result

$$\hat{s}(t) = \sum_n \frac{\mu_n}{\mu_n + N_0} r_n\phi_n(t). \tag{56}$$

The optimum estimator is then strictly linear.

B. Performance

The MSE in the interval $[0, T]$, obtained by using the WL estimator $\hat{\boldsymbol{\sigma}}(t)$ in (54), is

$$\begin{aligned}\text{MSE} &= \frac{1}{2}\int_0^T \operatorname{tr} E\left[\boldsymbol{\sigma}(t)(\boldsymbol{\sigma}(t) - \hat{\boldsymbol{\sigma}}(t))^{\mathsf{H}}\right] dt \\ &= \frac{1}{2}\int_0^T \operatorname{tr}\left[\boldsymbol{R}(t,t) - \int_0^T \boldsymbol{H}(t,v)\boldsymbol{R}(t,v)dv\right] dt \\ &= \frac{N_0}{2}\sum_n \left[\frac{\lambda_n^r}{\lambda_n^r + N_0} + \frac{\lambda_n^i}{\lambda_n^i + N_0}\right].\end{aligned} \tag{57}$$

According to Fact 2, the MSE is a Schur-concave function of the eigenvalues $\{\lambda_n^r, \lambda_n^i\}$ with parameter N_0, the noise level. Thus, Theorem 2 is applicable and to evaluate the maximum advantage of WL over SL processing, we consider the situation where the disparity between the two sets of eigenvalues is largest, for instance, $\lambda_n^r = 2\mu_n$ and $\lambda_n^i = 0$. Then the optimum choice for $h_{n,1}$ and $h_{n,2}$ is

$$h_{n,1} = h_{n,2} = \frac{\mu_n}{2\mu_n + N_0}. \tag{58}$$

This produces the estimator

$$\hat{s}(t) = \sum_n \frac{2\mu_n}{2\mu_n + N_0} \operatorname{Re}(r_n)(\phi_n(t) + \psi_n(t)) \tag{59}$$

which only depends on $\operatorname{Re} r_n$, since this is a sufficient statistic for this problem. The MSE is

$$\text{MSE} = \frac{N_0}{2} \sum_n \frac{2\mu_n}{2\mu_n + N_0}. \tag{60}$$

SL processing, on the other hand, ignores complementary covariances. Here, we cannot distinguish between the two sets of eigenvalues, and therefore we must assume that they are equal, i.e., $\lambda_n^r = \lambda_n^i = \mu_n$. This leads to the suboptimal coefficients

$$h'_{n,1} = \frac{\mu_n}{\mu_n + N_0}, \quad h'_{n,2} = 0. \tag{61}$$

The well-known formula for the MSE is then

$$\text{MSE}' = \frac{N_0}{2} \sum_n \frac{2\mu_n}{\mu_n + N_0}. \tag{62}$$

Therefore, as $N_0 \to 0$, WL processing promises half the MSE of SL processing. This result was obtained in [12] for the WSS case. The advantage of WL filtering diminishes as N_0 becomes larger and disappears completely for $N_0 \to \infty$. For the communications example in Section II, with measurements given by (23), the maximum advantage of coherent over noncoherent processing is the well-known 3-dB gain.

We make two important remarks. First, if the two sets of eigenvalues of the augmented covariance of $s(t)$ do not exhibit maximum disparity, then the maximum performance advantage, which will be less than 3 dB, occurs for some noise level $N_0 > 0$. Second, if the noise $n(t)$ is improper, the performance advantage of WL over SL processing can be arbitrarily large. To see this, consider the case where the signal $s(t)$ is real and the noise $n(t)$ is purely imaginary. In this case, the WL operation $\operatorname{Re} r(t)$ will yield a *perfect* estimate of $s(t)$. The SL estimator, on the other hand, will suffer from a nonzero estimation error. Hence, the relative advantage of WL over SL processing would be infinite in this special case.

V. Detection

The detection problem we would like to solve is a simple hypothesis test

$$\begin{aligned} H_0 &: r(t) = n(t) \\ H_1 &: r(t) = s(t) + n(t). \end{aligned} \tag{63}$$

We observe the complex signal $r(t)$ over the time interval $0 \leq t \leq T$. The noise $n(t)$ is zero-mean complex white (i.e., proper) Gaussian with power spectral density N_0, and the zero-mean complex Gaussian signal $s(t)$ is described by its augmented covariance $\boldsymbol{R}(t_1, t_2)$. This detection problem possesses well-known solutions in the real and proper cases [14], [15].

A. Solution of the Detection Problem

In the improper case, we first determine the K-L expansion of $s(t)$ as described in Theorem 1. The K-L coefficients s_n are improper complex Gaussian random variables with

$$E(s_n s_m^*) = \frac{1}{2}(\lambda_n^r + \lambda_n^i)\delta_{nm}$$

and

$$E(s_n s_m) = \frac{1}{2}(\lambda_n^r - \lambda_n^i)\delta_{nm}$$

and the resolution of the observed signal $r(t)$ onto the K-L basis functions is given by (53). Let

$$\begin{aligned} \boldsymbol{\Lambda}_r &= \operatorname{diag}[\lambda_1^r, \ldots, \lambda_N^r] \\ \boldsymbol{\Lambda}_i &= \operatorname{diag}[\lambda_1^i, \ldots, \lambda_N^i] \end{aligned}$$

and $\boldsymbol{r} = [r_1, \ldots, r_N]^{\mathsf{T}}$. The finite-dimensional log-likelihood ratio is then

$$\begin{aligned} L &= \frac{1}{2}\left[\boldsymbol{r}^{\mathsf{H}} \quad \boldsymbol{r}^{\mathsf{T}}\right] \left(\begin{bmatrix} N_0\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & N_0\boldsymbol{I} \end{bmatrix}^{-1} \right. \\ &\quad \left. - \begin{bmatrix} \frac{1}{2}(\boldsymbol{\Lambda}_r + \boldsymbol{\Lambda}_i) + N_0\boldsymbol{I} & \frac{1}{2}(\boldsymbol{\Lambda}_r - \boldsymbol{\Lambda}_i) \\ \frac{1}{2}(\boldsymbol{\Lambda}_r - \boldsymbol{\Lambda}_i) & \frac{1}{2}(\boldsymbol{\Lambda}_r + \boldsymbol{\Lambda}_i) + N_0\boldsymbol{I} \end{bmatrix}^{-1} \right) \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{r}^* \end{bmatrix} \\ &= \frac{1}{N_0} \sum_{n=1}^{N} h_{n,1}|r_n|^2 + h_{n,2} \operatorname{Re} r_n^2 \\ &= \frac{1}{N_0} \operatorname{Re}\left(\sum_{n=1}^{N} \hat{s}_n^* r_n \right) \end{aligned} \tag{64}$$

where $h_{n,1}$ and $h_{n,2}$ are given by (51) and (52), and $\hat{s}_n$ is the WL estimator of (55). Thanks to Grenander's theorem [14, Sec.VI.B.2], L converges to the log-likelihood ratio of the infinite-dimensional detection problem (63) as $N \to \infty$. Note that (64) shows that a detector for an improper signal can still be expressed as a cascade of an estimator and a correlator just like the detector for a proper signal. Moreover, the log-likelihood ratio (64) simplifies to the known result in the proper case:

$$L = \frac{1}{N_0} \sum_n \frac{\mu_n}{\mu_n + N_0} |r_n|^2. \tag{65}$$

B. Performance

To evaluate the performance of the detector, we use the deflection criterion [7], which is defined as

$$D(L) = \frac{(E_1(L) - E_0(L))^2}{V_0(L)}. \tag{66}$$

Here, $E_i(\cdot)$ denotes expectation under hypothesis i, and $V_0(\cdot)$ denotes variance under hypothesis zero. The deflection can be regarded as an output signal-to-noise ratio. It is much easier to compute than a complete receiver operator characteristic curve.

Computation of $D(L)$ is straightforward, except for the evaluation of $V_0(L) = E_0(L^2) - E_0^2(L)$, because $E_0(L^2)$ involves fourth-order properties of complex white Gaussian noise. The Gaussian assumption forces

$$\begin{aligned} E\left(n_i^{S_i} n_j^{S_j} n_k^{S_k} n_l^{S_l}\right) &= E\left(n_i^{S_i} n_j^{S_j}\right) E\left(n_k^{S_k} n_l^{S_l}\right) \\ &\quad + E\left(n_i^{S_i} n_k^{S_k}\right) E\left(n_j^{S_j} n_l^{S_l}\right) \\ &\quad + E\left(n_i^{S_i} n_l^{S_l}\right) E\left(n_j^{S_j} n_k^{S_k}\right) \end{aligned} \tag{67}$$

where S_m is either $+1$ or the conjugating star $*$. Since the noise is white and proper, we have $E n_n n_m^* = N_0 \delta_{nm}$ and $E n_n n_m = 0$, and the only nonzero fourth-order moment of the noise must have exactly two star and two nonstar terms. This moment is then

$$En_i n_j n_k^* n_l^* = N_0^2 \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right). \tag{68}$$

Using (64) we obtain

$$\begin{aligned} E_0(L^2) &= \frac{1}{N_0^2}\sum_n \sum_m E_0 (h_{n,1}h_{m,1}n_n n_m n_n^* n_m^* \\ &\qquad + \frac{1}{2} h_{n,2}h_{m,2} n_n n_n n_m^* n_m^*) \\ &= \left(\sum_n h_{n,1}\right)^2 + \sum_n (h_{n,1}^2 + h_{n,2}^2). \end{aligned} \tag{69}$$

The deflection is then given as

$$\begin{aligned} D(L) &= \frac{1}{4N_0^2}\cdot\frac{\left(\sum_n h_{n,1}(\lambda_n^r+\lambda_n^i) + h_{n,2}(\lambda_n^r-\lambda_n^i)\right)^2}{\sum_n (h_{n,1}^2+h_{n,2}^2)} \\ &= \frac{1}{2N_0^2}\cdot\frac{\left(\sum_n \frac{(\lambda_n^r)^2}{\lambda_n^r+N_0} + \frac{(\lambda_n^i)^2}{\lambda_n^i+N_0}\right)^2}{\sum_n \left[\left(\frac{\lambda_n^r}{\lambda_n^r+N_0}\right)^2 + \left(\frac{\lambda_n^i}{\lambda_n^i+N_0}\right)^2\right]}. \end{aligned} \tag{70}$$

Using Fact 1, we can show that $D(L)$ is a Schur-convex function of the eigenvalues $\{\lambda_n^r, \lambda_n^i\}$, with parameter N_0. The easy, but tedious, proof is omitted. To determine the maximum performance gain by incorporating knowledge of complementary covariances we can therefore again employ Theorem 2. In order to maximize deflection, the two eigenvalues λ_n^r and λ_n^i must show maximum disparity: for instance, $\lambda_n^r = 2\mu_n$, $\lambda_n^i = 0$. Using the optimum coefficients $h_{n,1}$ and $h_{n,2}$ from (58), the deflection is then

$$D(L) = \frac{2}{N_0^2}\cdot\frac{\left(\sum_n \frac{\mu_n^2}{2\mu_n+N_0}\right)^2}{\sum_n \left(\frac{\mu_n}{2\mu_n+N_0}\right)^2}. \tag{71}$$

On the other hand, if we were to ignore the information contained in the complementary covariance, we would have $\lambda_n^r = \lambda_n^i = \mu_n$, and the suboptimal coefficients (61) would only give deflection

$$D'(L) = \frac{1}{N_0^2}\cdot\frac{\left(\sum_n \frac{\mu_n^2}{\mu_n+N_0}\right)^2}{\sum_n \left(\frac{\mu_n}{\mu_n+N_0}\right)^2}. \tag{72}$$

The ratio $D(L)/D'(L)$ is at most 2, and this is asymptotically the case for either $N_0 \to 0$ or $N_0 \to \infty$. Again, this 3-dB result is not surprising for the comparison of coherent versus noncoherent detection in the communications example of Section II, (23). Note, however, that our result is more general than just a comparison of coherent and noncoherent detection. We have proved that the factor of 2 is also the maximum performance gain possible in *any* Gauss–Gauss detector that incorporates knowledge of complementary covariances.

VI. Conclusion

We have presented a version of the K-L expansion for improper complex random signals. We have used this K-L expansion to solve the problems of detection and estimation of improper random signals in additive white Gaussian noise, and presented a general result comparing the performance of widely linear with strictly linear processing. In particular, we have found that the maximum performance gain by incorporating knowledge of complementary covariances as measured by MSE and deflection is a factor of 2. In a communications example, we have linked the two situations that display this maximum performance difference to coherent detection (perfect phase knowledge) and noncoherent detection (no phase knowledge). WL processing, however, can cover any situation in between, where *some* phase knowledge is available. It applies, as well, to adaptive approximations of coherent detection (see, e.g., [18], [19]).

Acknowledgment

The authors wish to thank an anonymous reviewer for encouraging them to clarify that WL processing can produce infinite advantage over SL processing when additive noise is improper.

References

[1] F. D. Neeser and J. L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
[2] B. Picinbono and P. Bondon, “Second-order statistics of complex signals,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–419, Feb. 1997.
[3] P. J. Schreier and L. L. Scharf, “Stochastic time-frequency analysis using the analytic signal: Why the complementary distribution matters,” *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3071–3079, Dec. 2003.
[4] B. Picinbono and P. Chevalier, “Widely linear estimation with complex data,” *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.
[5] B. Picinbono, “Second-order complex random vectors and normal distributions,” *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.
[6] P. J. Schreier and L. L. Scharf, “Second-order analysis of improper complex random vectors and processes,” *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.
[7] B. Picinbono, “On deflection as a performance criterion in detection,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 3, pp. 1072–1081, Jul. 1995.
[8] A. van den Bos, “The multivariate complex normal distribution—a generalization,” *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 537–539, Mar. 1995.
[9] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[10] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[11] R. C. Thompson and S. Therianos, “Inequalities connecting the eigenvalues of a Hermitian matrix with the eigenvalues of complementary principal submatrices,” *Bull. Austral. Math. Soc.*, vol. 6, pp. 117–132, 1972.
[12] W. M. Brown and R. B. Crane, “Conjugate linear filtering,” *IEEE Trans. Inf. Theory*, vol. IT-15, no. 4, pp. 462–465, Jul. 1969.
[13] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[14] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1998.
[15] H. L. Van Trees, *Detection, Estimation, and Modulation Theory—Part III*. New York: Wiley, 1971.
[16] E. J. Kelly and W. L. Root, “A representation of vector-valued random processes,” MIT Lincoln Lab., Lexington, MA, Group Rep. 55-21, 1960.
[17] H. L. Van Trees, *Detection, Estimation, and Modulation Theory—Part I*. New York: Wiley, 1968.
[18] G. Gelli, L. Paura, and A. R. P. Ragozini, “Blind widely linear multiuser detection,” *IEEE Commun. Lett.*, vol. 4, no. 6, pp. 187–189, Jun. 2000.
[19] S. Buzzi, M. Lops, and A. M. Tulino, “A new family of MMSE multiuser receivers for interference suppression in DS/CDMA systems employing BPSK modulation,” *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 154–167, Jan. 2001.