Complex-Valued Random Vectors and Channels: Entropy, Divergence, and Capacity

Georg Tauböck, *Member, IEEE*

***Abstract*—Recent research has demonstrated significant achievable performance gains by exploiting circularity/noncircularity or properness/improperness of complex-valued signals. In this paper, we investigate the influence of these properties on important information theoretic quantities such as entropy, divergence, and capacity. We prove two maximum entropy theorems that strengthen previously known results. The proof of the first maximum entropy theorem is based on the so-called *circular analog* of a given complex-valued random vector. The introduction of the circular analog is additionally supported by a characterization theorem that employs a minimum Kullback–Leibler divergence criterion. In the proof of the second maximum entropy theorem, results about the second-order structure of complex-valued random vectors are exploited. Furthermore, we address the capacity of multiple-input multiple-output (MIMO) channels. Regardless of the specific distribution of the channel parameters (noise vector and channel matrix, if modeled as random), we show that the capacity-achieving input vector is circular for a broad range of MIMO channels (including coherent and noncoherent scenarios). Finally, we investigate the situation of an improper and Gaussian distributed noise vector. We compute both capacity and capacity-achieving input vector and show that improperness increases capacity, provided that the complementary covariance matrix is exploited. Otherwise, a capacity loss occurs, for which we derive an explicit expression.**

***Index Terms*—Capacity, circular, circular analog, differential entropy, improper, Kullback–Leibler divergence, multiple-input multiple-output (MIMO), mutual information, noncircular, proper.**

I. INTRODUCTION

COMPLEX-VALUED signals are central in many scientific fields including communications, array processing, acoustics and optics, oceanography and geophysics, machine learning, and biomedicine. In recent research—for a comprehensive overview see [4]—it has been shown that exploiting circularity/properness of complex-valued signals or lack of it (noncircularity/improperness) is able to significantly enhance the performance of signal processing techniques. More specifically, in the field of communications, it has been observed that important digital modulation schemes including binary phase-shift keying, pulse amplitude modulation, Gaussian minimum shift keying, offset quaternary phase-shift keying, and *baseband* (but not passband) orthogonal frequency division multiplexing—commonly called discrete multitone (DMT)—produce noncircular/improper complex baseband signals under various circumstances (see e.g., [5]–[10]). Noncircular/improper baseband communication signals can also arise due to imbalance between their in-phase and quadrature (I/Q) components, and several techniques for compensating for I/Q imbalance have been proposed [11]–[13].

Information theory, on the other hand, addresses fundamental performance limits of communication systems and also has large impact on many other scientific areas, where stochastic models are used. Clearly, it has the potential to study performance limits of signal processing algorithms and communication systems that exploit circularity/properness or noncircularity/improperness. However, there exist only a few information theoretic results addressing these questions [4], [14] and further advancements are desirable. A significant disadvantage of available results is that they often stick to a Gaussian assumption, something which is not always the case in practice.

Apparently the first information theoretic work that addresses complex-valued random vectors analyzes the differential entropy of complex-valued random vectors [14]. More specifically, it is shown that the differential entropy of a zero-mean complex-valued random vector with given covariance matrix is upper bounded by the differential entropy of a circular (and, consequently, zero mean and proper) Gaussian distributed complex-valued random vector with the same covariance matrix. Using this result, capacity results for complex-valued multiple-input multiple-output (MIMO) channels with additive circular/proper Gaussian noise have been derived [15]. In particular, it has been shown that the capacity-achieving input vector is Gaussian distributed and circular/proper.

Let us suppose that we are dealing with a complex-valued random vector which is known to be non-Gaussian. In this situation, the differential entropy upper bound developed in [14] turns out to be not tight. The same is the case, if the complex-valued random vector is known to be improper. Hence, there are two sources that decrease the differential entropy of a complex-valued random vector, i.e., non-Gaussianity and improperness. One goal of this paper is to derive improved/tighter maximum entropy theorems for both cases.

The maximum entropy theorem in [14] associates a circular random vector (i.e., the Gaussian distributed one) to a given

Manuscript received August 09, 2006; revised October 31, 2010; accepted May 25, 2011. Date of publication January 31, 2012; date of current version April 17, 2012. This work was supported by WWTF Grant MOHAWI (MA 44) and WWTF Grant SPORTS (MA 07-004) and by FWF Grant "Statistical Inference" (S10603-N13) within the National Research Network SISE. The material in this paper was presented in part at the 23rd Symposium on Information Theory in the Benelux, Louvain-la-Neuve, Belgium, May 2002; at the 2004 IEEE International Symposium on Information Theory, and at the 2011 IEEE Information Theory Workshop [1]–[3].

The author is with the Institute of Telecommunications, Vienna University of Technology, Gusshausstrasse 25/389, 1040 Vienna, Austria (e-mail: gtauboec@nt.tuwien.ac.at).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2184638

complex-valued random vector. As pointed out, this choice does not always lead to the smallest change in differential entropy. This raises the question, how we can associate a circular random vector to an (in general) non-circular complex-valued random vector in a canonical way but not forcing it to be Gaussian distributed. The choice we propose is intuitive and is, furthermore, supported by a characterization theorem that is based on a minimum Kullback–Leibler divergence criterion. It also leads to the desired improved entropy upper bound for the case for which the random vector is known to be non-Gaussian. A study of further properties complements its analysis.

As already mentioned, the maximum entropy theorem in [14] does not yield a tight upper bound if the random vector is known to be improper. Extending our work of [1] and [2], we derive an improved maximum entropy theorem which addresses this situation. As a by-product, we obtain a criterion for a matrix to be a valid complementary covariance matrix (also termed pseudo-covariance matrix [14]). We note that after our initial work [1], [2], the obtained characterization of complementary covariance matrices has been extended in [16] and [17]. Meanwhile, expressions for the differential entropy of improper Gaussian random vectors have appeared in the literature as well [4], [18].

Finally, we apply the obtained maximum entropy theorems to derive novel capacity results for complex-valued MIMO channels with additive noise vectors. Without making use of any Gaussian assumption (in contrast to [15]), we show that capacity is achieved by circular random vectors for a broad range of channels. These results include both the case of a deterministic channel matrix and the case of a random channel matrix, whose realizations are assumed to be either known to the receiver (coherent capacity) or unknown (incoherent capacity). On the other hand, we investigate the capacity of channels, whose noise is noncircular/improper and Gaussian distributed. Such channels have been shown to occur if, e.g., DMT is used as modulation scheme [9], [10]. Note that DMT is currently employed in several *x*DSL standards [19]. We derive capacity expressions for two cases: 1) we assume that the knowledge of the complementary covariance matrix is taken into account (both at transmitter and receiver); and 2) we assume that it is erroneously believed—i.e., that the transceiver is designed assuming—that the noise has a vanishing complementary covariance matrix, so that the information contained in the complementary covariance matrix is ignored. This results in a decreased capacity and we calculate the occurring capacity loss.

Notation: The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$. We use the superscript $[\cdot]^T$ for transposition and the superscript $[\cdot]^H \triangleq \left([\cdot]^T\right)^*$ for Hermitian transposition, where the superscript $[\cdot]^*$ stands for complex conjugation. $j = \sqrt{-1}$ denotes the imaginary unit, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are real and imaginary parts, $\mathrm{E}\{\cdot\}$ refers to usual expectation, and $\det(\cdot)$ and $\mathrm{tr}(\cdot)$ express determinant and trace of a matrix, respectively. Throughout the paper, $\log(\cdot)$ denotes the logarithm taken with respect to an arbitrary but fixed base. Therefore, all results are valid regardless of the chosen unit for differential entropy (*nats* or *bits*).

Outline: The remainder of this paper is organized as follows. In Section II, we introduce our framework and present initial results about the distribution and second-order properties of complex-valued random vectors. Section III deals with the question, how to circularize complex-valued random vectors and analyzes the proposed method. The differential entropy of complex-valued random vectors is addressed in Section IV and two improved maximum entropy theorems are proved. Finally, in Section V, we present various capacity results for complex-valued MIMO channels.

II. Framework and Preliminary Results

We consider complex-valued random vectors $\mathbf{x} \in \mathbb{C}^n$. We assume that $\mathbf{x}^{(\mathrm{r})} \in \mathbb{R}^{2n}$, where $\mathbf{x}^{(\mathrm{r})} \triangleq \left[\Re\{\mathbf{x}^T\}\ \Im\{\mathbf{x}^T\}\right]^T$ is defined by stacking of real and imaginary part of $\mathbf{x}$, is distributed according to a joint multivariate $2n$-dimensional probability density function (pdf) $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})$. More precisely, it is assumed that the measure[1] defining the distribution of $\mathbf{x}^{(\mathrm{r})}$ is absolutely continuous with respect to λ_{2n}, where λ_{2n} denotes the $2n$-dimensional Lebesgue measure [20]. Accordingly, whenever an integral appears in this paper, integration is meant with respect to the Lebesgue measure of appropriate dimension. Note that when we refer to the distribution of $\mathbf{x}$, we mean the distribution of $\mathbf{x}^{(\mathrm{r})}$ defined by the pdf $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})$. Hence, a complex-valued random vector $\mathbf{x}$ will be called Gaussian distributed if $\mathbf{x}^{(\mathrm{r})}$ is (multivariate) Gaussian distributed.

Definition 2.1: A complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is said to be *circular*, if $\mathbf{x}$ has the same distribution as $e^{j2\pi\theta}\mathbf{x}$ for all $\theta \in [0,1)$; otherwise it is said to be *noncircular*. The set of all circular complex-valued random vectors $\mathbf{x} \in \mathbb{C}^n$, whose distribution is absolutely continuous with respect to λ_{2n}, is denoted by $\mathcal{C}_n$.

It is well known (see, e.g., [1], [4], [14], [21]) that for a complete second-order characterization of a complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ not only mean vector $\mathbf{m_x} \triangleq \mathrm{E}\{\mathbf{x}\}$ and covariance matrix $\mathbf{C_x} \triangleq \mathrm{E}\left\{(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^H\right\}$ but also *complementary covariance matrix* $\mathbf{P_x} \triangleq \mathrm{E}\left\{(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^T\right\}$ are required. Note that both mean vector and complementary covariance matrix of a circular complex-valued random vector are vanishing provided that its first- and second-order moments exist [4].

Definition 2.2: A complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is said to be *proper*, if its complementary covariance matrix vanishes; otherwise, it is said to be *improper*.

Hence, circularity implies properness (under the assumption of existing first- and second-order moments). Note that a zero-mean and proper Gaussian random vector is circular.

A. Polar and Sheared-Polar Representation

Here, we present some auxiliary results about the distribution of complex-valued random vectors. Let us denote by $\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}$ the mapping

$$\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}: \begin{cases} (\mathbb{R}_0^+)^n \times [0,1)^n \to \mathbb{R}^{2n}, \\ [r_1 \cdots r_n\ \phi_1 \cdots \phi_n]^T \mapsto [r_1 \cos(2\pi\phi_1) \cdots \\ \quad r_n \cos(2\pi\phi_n)\ \ r_1 \sin(2\pi\phi_1) \cdots r_n \sin(2\pi\phi_n)]^T \end{cases}$$

[1]Here, we refer to the measure defined on the Borel σ-field on $\mathbb{R}^{2n}$ induced by the measurable function defining the random vector.

where $\mathbb{R}_0^+$ denotes the set of nonnegative reals. There exists the inverse $\mathbb{T}^{(\mathrm{r}\to\mathrm{p})} \triangleq \left(\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}\right)^{-1}$, provided that we set $\phi_i \triangleq 0$ for $r_i = 0$, $i = 1, \ldots, n$. Note that the set, by which the domain of $\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}$ is reduced according to this convention has measure zero with respect to λ_{2n}. In the following, $\mathbf{x}^{(\mathrm{r})}$ will be called *real representation* of $\mathbf{x}$, whereas $\mathbf{x}^{(\mathrm{p})} \triangleq \mathbb{T}^{(\mathrm{r}\to\mathrm{p})}(\mathbf{x}^{(\mathrm{r})})$ will be denoted as *polar representation* of $\mathbf{x}$.

Lemma 2.3: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector, which is distributed according to the pdf $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})$. Then, the pdf of its polar representation $\mathbf{x}^{(\mathrm{p})}$ is given by

$$f_{\mathbf{x}^{(\mathrm{p})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) = \begin{cases} (2\pi)^n (r_1 \cdots r_n) f_{\mathbf{x}^{(\mathrm{r})}}\big(\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n)\big), \\ \qquad (r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) \in (\mathbb{R}_0^+)^n \times [0,1)^n \\ 0, \qquad \text{otherwise} \end{cases}$$

almost everywhere with respect to λ_{2n} (λ_{2n}-a.e.) [20].

Proof: Follows from

$$\int_{\mathcal{A}} f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_{\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(\mathcal{A})} f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_{\mathcal{A}} f_{\mathbf{x}^{(\mathrm{r})}}\left(\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(\boldsymbol{\xi})\right) \left|J_{\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}}(\boldsymbol{\xi})\right| d\boldsymbol{\xi}$$

for all Lebesgue measurable sets $\mathcal{A} \subset (\mathbb{R}_0^+)^n \times [0,1)^n$, where the Jacobian determinant $J_{\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}}(\boldsymbol{\xi})$ of $\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}$ is easily computed as $J_{\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) = (2\pi)^n (r_1 \cdots r_n)$ [22]. □

Observing that the pdf of $\mathbf{y}_{(\theta)}^{(\mathrm{p})}$ of the random vector $\mathbf{y}_{(\theta)} \triangleq e^{j2\pi\theta}\mathbf{x}$ (with $\theta \in [0,1)$ being deterministic) satisfies $f_{\mathbf{y}_{(\theta)}^{(\mathrm{p})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) = f_{\mathbf{x}^{(\mathrm{p})}}\big(r_1, \ldots, r_n, [\phi_1 - \theta]_{[0,1)}, \ldots, [\phi_n - \theta]_{[0,1)}\big)$ λ_{2n}-a.e., where the notation $[\cdot]_{[0,1)}$ is shorthand for modulo with respect to the interval $[0,1)$, we obtain the following corollary.

Corollary 2.4: A complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is circular if and only if the pdf of its *polar representation* $\mathbf{x}^{(\mathrm{p})}$ satisfies

$$f_{\mathbf{x}^{(\mathrm{p})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) = f_{\mathbf{x}^{(\mathrm{p})}}\big(r_1, \ldots, r_n, [\phi_1 - \theta]_{[0,1)}, \ldots, [\phi_n - \theta]_{[0,1)}\big) \quad \forall \theta \in [0,1) \quad \lambda_{2n}\text{-a.e.}.$$

Let us denote by $\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}$ the mapping shown in the first equation at the bottom of the page, which is one-to-one with inverse $\mathbb{T}^{(\mathrm{p}\to\mathrm{s})} \triangleq \left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}\right)^{-1}$ given by the second equation shown at the bottom of the page. This follows immediately from the identity

$$[\phi]_{[0,1)} = \phi + n(\phi), \quad \phi \in \mathbb{R} \tag{1}$$

where $n(\phi) \in \mathbb{Z}$. In the following, $\mathbf{x}^{(\mathrm{s})} \triangleq \mathbb{T}^{(\mathrm{p}\to\mathrm{s})}(\mathbf{x}^{(\mathrm{p})}) = \mathbb{T}^{(\mathrm{p}\to\mathrm{s})}\left(\mathbb{T}^{(\mathrm{r}\to\mathrm{p})}(\mathbf{x}^{(\mathrm{r})})\right)$ will be called *sheared-polar representation* of $\mathbf{x}$.

Lemma 2.5: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the pdfs of its polar representation $\mathbf{x}^{(\mathrm{p})}$ and its sheared-polar representation $\mathbf{x}^{(\mathrm{s})}$ are related according to

$$f_{\mathbf{x}^{(\mathrm{s})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) = f_{\mathbf{x}^{(\mathrm{p})}}\big(r_1, \ldots, r_n, [\phi_1 + \phi_n]_{[0,1)}, \ldots, [\phi_{n-1} + \phi_n]_{[0,1)}, \phi_n\big) \quad \lambda_{2n}\text{-a.e.}$$

$$f_{\mathbf{x}^{(\mathrm{p})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) = f_{\mathbf{x}^{(\mathrm{s})}}\big(r_1, \ldots, r_n, [\phi_1 - \phi_n]_{[0,1)}, \ldots, [\phi_{n-1} - \phi_n]_{[0,1)}, \phi_n\big) \quad \lambda_{2n}\text{-a.e.}.$$

Proof: Observe that the measure defining the distribution of $\mathbf{x}^{(\mathrm{s})}$ is absolutely continuous with respect to λ_{2n}, since $\lambda_{2n}(\mathcal{N}) = 0$ implies $\lambda_{2n}(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{N})) = 0$ for all $\mathcal{N} \subset (\mathbb{R}_0^+)^n \times [0,1)^n$, as can be seen by distinction of cases according to the modulo-$[0,1)$ operation. Since $\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}$ is not continuous on its whole domain, we define the auxiliary mapping $\tilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})} : (\mathbb{R}_0^+)^n \times \mathbb{R}^n \to (\mathbb{R}_0^+)^n \times \mathbb{R}^n$, $[r_1 \cdots r_n \ \phi_1 \cdots \phi_n]^T \mapsto [r_1 \cdots r_n \ (\phi_1 + \phi_n) \cdots (\phi_{n-1} + \phi_n) \ \phi_n]^T$, which is one-to-one with inverse $\tilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})} \triangleq \left(\tilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}\right)^{-1}$, where $\tilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})} : (\mathbb{R}_0^+)^n \times \mathbb{R}^n \to (\mathbb{R}_0^+)^n \times \mathbb{R}^n$, $[r_1 \cdots r_n \ \phi_1 \cdots \phi_n]^T \mapsto [r_1 \cdots r_n \ (\phi_1 - \phi_n) \cdots (\phi_{n-1} - \phi_n) \ \phi_n]^T$. Its Jacobian determinant is identically $J_{\tilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}}(r_1, \ldots, r_n, \phi_1, \ldots, \phi_n) \equiv 1$. Suppose $\mathcal{A} \subset (\mathbb{R}_0^+)^n \times [0,1)^n$ is any set of the form $\mathcal{A} = [a_1, b_1) \times \cdots \times [a_{2n}, b_{2n})$. From (1), it follows that there exists a finite partition $\{\mathcal{A}_1, \ldots, \mathcal{A}_N\}$ of $\mathcal{A}$, i.e., $\mathcal{A} = \bigcup_{i=1}^{N} \mathcal{A}_i$ and $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ for $i \neq j$, such that $\tilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A}_i)\right) = \mathcal{A}_i + \mathbf{k}_i$ and $\tilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A})\right) = \bigcup_{i=1}^{N} (\mathcal{A}_i + \mathbf{k}_i)$, where $\mathbf{k}_i \in \{0\}^n \times \mathbb{Z}^{n-1} \times \{0\}$. Here, $\mathcal{A}_i + \mathbf{k}_i$ denote the disjoint sets $\mathcal{A}_i + \mathbf{k}_i \triangleq \{\boldsymbol{\xi} \in \mathbb{R}^{2n} : \boldsymbol{\xi} - \mathbf{k}_i \in \mathcal{A}_i\}$, $i = 1, \ldots, N$. Note that the partition is caused by the modulo-$[0,1)$ operation used in the

---

$$\mathbb{T}^{(\mathrm{s}\to\mathrm{p})} : \begin{cases} (\mathbb{R}_0^+)^n \times [0,1)^n & \to \quad (\mathbb{R}_0^+)^n \times [0,1)^n, \\ [r_1 \cdots r_n \ \phi_1 \cdots \phi_n]^T & \mapsto \quad \left[r_1 \cdots r_n \, [\phi_1 + \phi_n]_{[0,1)} \cdots [\phi_{n-1} + \phi_n]_{[0,1)} \ \phi_n\right]^T \end{cases}$$

$$\mathbb{T}^{(\mathrm{p}\to\mathrm{s})} : \begin{cases} (\mathbb{R}_0^+)^n \times [0,1)^n & \to \quad (\mathbb{R}_0^+)^n \times [0,1)^n, \\ [r_1 \cdots r_n \ \phi_1 \cdots \phi_n]^T & \mapsto \quad \left[r_1 \cdots r_n \, [\phi_1 - \phi_n]_{[0,1)} \cdots [\phi_{n-1} - \phi_n]_{[0,1)} \ \phi_n\right]^T \end{cases}$$

definition of $\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}$ and corresponds to the required distinction of cases when investigating $\widetilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A})\right)$. Therefore

$$\begin{aligned}\int_{\mathcal{A}} f_{\mathbf{x}^{(\mathrm{s})}}(\boldsymbol{\xi})d\boldsymbol{\xi} &= \int_{\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A})} f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= \int_{\widetilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A})\right)} f_{\mathbf{x}^{(\mathrm{p})}}\left(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\right)\left|J_{\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}}(\boldsymbol{\xi})\right| d\boldsymbol{\xi} \\ &= \bigcup_{i=1}^{N}\int_{\mathcal{A}_i+\mathbf{k}_i} f_{\mathbf{x}^{(\mathrm{p})}}\left(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\right)d\boldsymbol{\xi} \\ &= \bigcup_{i=1}^{N}\int_{\mathcal{A}_i} f_{\mathbf{x}^{(\mathrm{p})}}\left(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi}+\mathbf{k}_i)\right)d\boldsymbol{\xi} \\ &\overset{(*)}{=} \bigcup_{i=1}^{N}\int_{\mathcal{A}_i} f_{\mathbf{x}^{(\mathrm{p})}}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\right)d\boldsymbol{\xi} \\ &= \int_{\mathcal{A}} f_{\mathbf{x}^{(\mathrm{p})}}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\right)d\boldsymbol{\xi}\end{aligned}$$

where $(*)$ follows from (1) and the fact that $\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi}+\mathbf{k}_i) \in (\mathbb{R}_0^+)^n \times [0,1)^n$ for $\boldsymbol{\xi} \in \mathcal{A}_i$. This implies the statement (see, e.g., [20]). □

Combining Corollary 2.4 and Lemma 2.5, while applying (1), yields the following important corollary.

Corollary 2.6: A complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is circular if and only if the pdf of its *sheared-polar representation* $\mathbf{x}^{(\mathrm{s})}$ does not depend on ϕ_n, i.e., if and only if

$$f_{\mathbf{x}^{(\mathrm{s})}}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n)=f_{\mathbf{x}^{(\mathrm{s})}}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_{n-1}) \quad \lambda_{2n}\text{-a.e.}\ .$$

B. Second-Order Properties

In the following, we establish some results about covariance and complementary covariance matrices of complex-valued random vectors. For a given complex-valued matrix $\mathbf{A} \in \mathbb{C}^{n\times m}$, let us denote by $\bar{\mathbf{A}} \in \mathbb{R}^{2n\times 2m}$ and $\underline{\mathbf{A}} \in \mathbb{R}^{2n\times 2m}$ the real-valued matrices

$$\overline{\mathbf{A}} \triangleq \begin{bmatrix}\Re\{\mathbf{A}\} & -\Im\{\mathbf{A}\}\\ \Im\{\mathbf{A}\} & \Re\{\mathbf{A}\}\end{bmatrix} \quad \text{and} \tag{2a}$$

$$\underline{\mathbf{A}} \triangleq \begin{bmatrix}\Re\{\mathbf{A}\} & \Im\{\mathbf{A}\}\\ \Im\{\mathbf{A}\} & -\Re\{\mathbf{A}\}\end{bmatrix} \tag{2b}$$

respectively. This notation allows a simple expression of the covariance matrix of the real representation $\mathbf{C}_{\mathbf{x}^{(\mathrm{r})}}$ of a complex-valued random vector $\mathbf{x}$ in terms of covariance matrix $\mathbf{C}_{\mathbf{x}}$ and complementary covariance matrix $\mathbf{P}_{\mathbf{x}}$ as [1], [4], [14], [21]

$$\mathbf{C}_{\mathbf{x}^{(\mathrm{r})}} = \frac{1}{2}\overline{\mathbf{C}}_{\mathbf{x}} + \frac{1}{2}\underline{\mathbf{P}}_{\mathbf{x}}. \tag{3}$$

Furthermore, $\mathbf{A}$, $\overline{\mathbf{A}}$, and $\underline{\mathbf{A}}$ satisfy remarkable algebraic properties, as stated by the next lemma.

Lemma 2.7:

$$\mathbf{C} = \mathbf{A}\mathbf{B} \iff \overline{\mathbf{C}} = \overline{\mathbf{A}}\,\overline{\mathbf{B}} \iff \underline{\mathbf{C}} = \overline{\mathbf{A}}\underline{\mathbf{B}} \tag{4a}$$

$$\mathbf{C} = \mathbf{A}\mathbf{B}^* \iff \underline{\mathbf{C}} = \underline{\mathbf{A}}\overline{\mathbf{B}} \tag{4b}$$

$$\mathbf{C} = \mathbf{A}^H \iff \overline{\mathbf{C}} = \overline{\mathbf{A}}^T \tag{4c}$$

$$\mathbf{U} \in \mathbb{C}^{n\times n}\text{unitary} \iff \overline{\mathbf{U}} \in \mathbb{R}^{2n\times 2n}\text{orthonormal} \tag{4d}$$

$$\det\overline{\mathbf{A}} = |\det\mathbf{A}|^2 = \det\left(\mathbf{A}\mathbf{A}^H\right), \quad \mathbf{A} \in \mathbb{C}^{n\times n}. \tag{4e}$$

Proof: For some of the statements, see also [15]. Direct calculations yield (4a) and (4b). Equation (4c) follows from the definition of $\overline{\mathbf{A}}$. A combination of (4a) and (4c), while observing that $\overline{\mathbf{I}}_n = \mathbf{I}_{2n}$, yields (4d). Finally, for (4e)

$$\begin{aligned}\det\overline{\mathbf{A}} &= \det\left(\begin{bmatrix}\mathbf{I}_n & j\mathbf{I}_n\\ \mathbf{0} & \mathbf{I}_n\end{bmatrix}\overline{\mathbf{A}}\begin{bmatrix}\mathbf{I}_n & -j\mathbf{I}_n\\ \mathbf{0} & \mathbf{I}_n\end{bmatrix}\right) \\ &= \det\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \Im\{\mathbf{A}\} & \mathbf{A}^*\end{bmatrix} \\ &= \det\mathbf{A}\det\mathbf{A}^*.\end{aligned}$$

□

We are especially interested in the eigenvalues of $\underline{\mathbf{P}}_{\mathbf{x}}$. We will show that they are essentially given by the *singular values* of $\mathbf{P}_{\mathbf{x}}$. Note that the *singular value decomposition* (SVD) [23] of a matrix $\mathbf{A} \in \mathbb{C}^{n\times m}$ factorizes $\mathbf{A}$ into three matrices, i.e., $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$. It is well defined for all rectangular complex matrices and yields unitary matrices $\mathbf{U} \in \mathbb{C}^{n\times n}$ and $\mathbf{V} \in \mathbb{C}^{m\times m}$ and a diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times m}$, i.e., $\boldsymbol{\Lambda} = \mathrm{diag}^{n\times m}\left\{\lambda_1,\ldots,\lambda_{\min\{n,m\}}\right\}$, with nonnegative entries on its main diagonal—the singular values. $\mathbf{U}$ and $\mathbf{V}$ can be chosen such that the singular values are ordered in decreasing order. In case the matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is symmetric (not Hermitian), i.e., $\mathbf{A}^T = \mathbf{A}$, there is a special SVD known as *Takagi factorization* [24]. It is given by the factorization

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T \tag{5}$$

where the columns of $\mathbf{Q}$ are the orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^{\mathbf{H}}$ and the diagonal matrix $\boldsymbol{\Lambda}$ has the singular values of $\mathbf{A}$ on its main diagonal.

Proposition 2.8: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector with complementary covariance matrix $\mathbf{P}_{\mathbf{x}} \in \mathbb{C}^{n\times n}$. Then, there exist a unitary matrix $\mathbf{Q}_{\mathbf{x}} \in \mathbb{C}^{n\times n}$ and a diagonal matrix $\boldsymbol{\Lambda}_{\mathbf{x}} \in \mathbb{R}^{n\times n}$ with nonnegative entries, such that

$$\underline{\mathbf{P}}_{\mathbf{x}} = \overline{\mathbf{Q}}_{\mathbf{x}}\underline{\boldsymbol{\Lambda}}_{\mathbf{x}}\overline{\mathbf{Q}}_{\mathbf{x}}^T$$

represents the eigenvalue decomposition of $\underline{\mathbf{P}}_{\mathbf{x}}$. The diagonal entries of $\boldsymbol{\Lambda}_{\mathbf{x}}$ are the singular values of $\mathbf{P}_{\mathbf{x}}$. In particular, $\underline{\boldsymbol{\Lambda}}_{\mathbf{x}} = \mathrm{diag}^{2n\times 2n}\{\boldsymbol{\Lambda}_{\mathbf{x}}, -\boldsymbol{\Lambda}_{\mathbf{x}}\}$.

Proof: Consider the Takagi factorization (5) of the symmetric $\mathbf{P}_{\mathbf{x}}$ and apply Lemma 2.7, i.e.,

$$\begin{aligned}\underline{\mathbf{P}}_{\mathbf{x}} &= \underline{\mathbf{Q}_{\mathbf{x}}\left(\boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{Q}_{\mathbf{x}}^T\right)} \\ &\overset{(4a)}{=} \overline{\mathbf{Q}}_{\mathbf{x}}\underline{\boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{Q}_{\mathbf{x}}^T}\end{aligned}$$

$$= \overline{\mathbf{Q}}_{\mathbf{x}} \mathit{\Lambda}_{\mathbf{x}} \underline{\left(\mathbf{Q}_{\mathbf{x}}^{H}\right)^{*}}$$
$$\stackrel{(4b)}{=} \overline{\mathbf{Q}}_{\mathbf{x}} \underline{\mathit{\Lambda}}_{\mathbf{x}} \overline{\mathbf{Q}_{\mathbf{x}}^{H}}$$
$$\stackrel{(4c)}{=} \overline{\mathbf{Q}}_{\mathbf{x}} \underline{\mathit{\Lambda}}_{\mathbf{x}} \overline{\mathbf{Q}}_{\mathbf{x}}^{T}$$

which represents the eigenvalue decomposition of $\underline{\mathbf{P}}_{\mathbf{x}}$, since $\overline{\mathbf{Q}}_{\mathbf{x}}$ is orthonormal according to (4d). □

An interesting question that is directly related to complex-valued random vectors is the characterization of the set of complementary covariance matrices. For covariance matrices such a characterization is well known, i.e., a matrix is a *valid* covariance matrix, which means that there exists a random vector with this covariance matrix, if and only if it is Hermitian and nonnegative definite. In order to obtain an analogous result for complementary covariance matrices, we introduce the following notion.[2]

Definition 2.9: A matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$ is said to be *generalized Cholesky factor* of a positive definite Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, if it satisfies $\mathbf{A} = \mathbf{B}\mathbf{B}^{H}$.

Since $\det \mathbf{A} = |\det \mathbf{B}|^2$, a generalized Cholesky factor is always a nonsingular matrix. Note that the conventional Cholesky decomposition (cf., [23]), $\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{H}}$, where $\mathbf{L}$ is lower-triangular, yields a generalized Cholesky factor $\mathbf{L}$. But there are also other ways of constructing a generalized Cholesky factor. Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathbf{H}}$ be the eigenvalue decomposition of $\mathbf{A}$. For any matrix $\mathbf{T}$, which satisfies $\mathbf{D} = \mathbf{T}\mathbf{T}^{\mathbf{H}}$, $\mathbf{B} = \mathbf{U}\mathbf{T}$ is a generalized Cholesky factor. Hence, a generalized Cholesky factor is not uniquely defined. However, we have the following characterization.

Proposition 2.10: Suppose $\mathbf{B}$ is a generalized Cholesky factor of $\mathbf{A}$. Then, for any unitary matrix $\mathbf{U}$, $\mathbf{C} = \mathbf{B}\mathbf{U}$ is also a generalized Cholesky factor. Conversely, if $\mathbf{B}$ and $\mathbf{C}$ are generalized Cholesky factors, then, there exists a unitary matrix $\mathbf{U}$, such that $\mathbf{C} = \mathbf{B}\mathbf{U}$.

Proof: For nonsingular $\mathbf{B}$ and $\mathbf{C}$, we have

$$\mathbf{B}\mathbf{B}^{H} = \mathbf{C}\mathbf{C}^{H} \iff \left(\mathbf{B}^{-1}\mathbf{C}\right)^{-1} = \left(\mathbf{B}^{-1}\mathbf{C}\right)^{H}$$

which implies both statements. □

The next theorem presents the promised criterion for a matrix to be a complementary covariance matrix. More precisely, it is a criterion in terms of both covariance matrix and complementary covariance matrix. We will call $\{\mathbf{C}, \mathbf{P}\}$ a *valid pair of covariance matrix and complementary covariance matrix*, if there exists a complex-valued random vector with covariance matrix $\mathbf{C}$ and complementary covariance matrix $\mathbf{P}$.

Theorem 2.11: Suppose $\mathbf{C} \in \mathbb{C}^{n\times n}$ is nonsingular and $\mathbf{P} \in \mathbb{C}^{n\times n}$. Then, $\{\mathbf{C}, \mathbf{P}\}$ is a valid pair of covariance matrix and complementary covariance matrix if and only if $\mathbf{C}$ is Hermitian and nonnegative definite, $\mathbf{P}$ is symmetric, and the singular values of $\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$ are smaller or equal to 1, where $\mathbf{B}$ denotes an arbitrary generalized Cholesky factor of $\mathbf{C}$.

[2]Cf. the relation to the *Karhunen-Loève transform* [25], [26], also known as *Hotelling transform* [27], and the *Mahalanobis transform*, e.g., [28] and references therein.

Proof: The requirements that $\mathbf{C}$ is Hermitian and nonnegative definite as well as that $\mathbf{P}$ is symmetric are obvious. Furthermore, observe that the singular values of $\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$ do not depend on the choice of the generalized Cholesky factor $\mathbf{B}$.

Suppose we are given a complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ with covariance matrix $\mathbf{C}$ and complementary covariance matrix $\mathbf{P}$. Consider the random vector $\mathbf{y} \stackrel{\triangle}{=} \mathbf{B}^{-1}\mathbf{x}$. Clearly, $\mathbf{C}_{\mathbf{y}} = \mathbf{I}_n$ and $\mathbf{P}_{\mathbf{y}} = \mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$. From (3), i.e., $\mathbf{C}_{\mathbf{y}^{(\mathrm{r})}} = \frac{1}{2}\left(\mathbf{I}_{2n} + \underline{\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}}\right)$, and the fact that $\mathbf{C}_{\mathbf{y}^{(\mathrm{r})}}$ is nonnegative definite, we conclude with Proposition 2.8, that the singular values of $\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$ are smaller or equal to 1.

Conversely, consider a complex-valued random vector $\mathbf{y}$, e.g., a Gaussian distributed one, defined by the covariance matrix of its real representation as $\mathbf{C}_{\mathbf{y}^{(\mathrm{r})}} \stackrel{\triangle}{=} \frac{1}{2}\left(\mathbf{I}_{2n} + \underline{\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}}\right)$. According to Proposition 2.8, such a random vector exists, since $\mathbf{C}_{\mathbf{y}^{(\mathrm{r})}}$ is Hermitian and nonnegative definite provided that the singular values of $\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$ are smaller or equal to 1. It has covariance matrix $\mathbf{C}_{\mathbf{y}} = \mathbf{I}_n$ and complementary covariance matrix $\mathbf{P}_{\mathbf{y}} = \mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$, cf., (3). Then, the random vector $\mathbf{x} \stackrel{\triangle}{=} \mathbf{B}\mathbf{y}$ has covariance matrix $\mathbf{C}_{\mathbf{y}} = \mathbf{C}$ and complementary covariance matrix $\mathbf{P}_{\mathbf{y}} = \mathbf{P}$. □

Remarks: Apparently, the importance of the singular values of $\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$ in the context of complex-valued random vectors was first observed in [1] (for the above criterion and a generalized maximum entropy theorem) and independently in [29], where they were introduced as *canonical coordinates* [17], [27], [30]–[32] between a complex-valued random vector and its complex conjugate. Note that in [29], the matrix $\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}$ is called *coherence matrix* between a random vector and its complex conjugate. Interestingly, the approach of Schreier and Scharf [29] differs from the approach taken here (and taken in [1]) in that [29] employs a complex-valued *augmented algebra* to study second-order properties of complex-valued random vectors, whereas (2) introduces a real-valued representation into real and imaginary parts. Later, in [18], the singular values were also termed *circularity coefficients* and the whole set of singular values was referred to as *circularity spectrum*. We also note that the condition of Theorem 2.11 on the singular values, can be equivalently expressed in terms of the Euclidean operator norm $\|\cdot\|_2$ as $\left\|\mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-T}\right\|_2 \leq 1$.

III. Circular Analog of a Complex-Valued Random Vector

In this section, we consider the following problem: suppose we are given a complex-valued random vector, which is noncircular. Can we find a random vector, which is as "similar" as possible to the original random vector but circular instead? Obviously, this depends on what is meant by "similar" and is, therefore, mainly a matter of definition. However, if we can show useful properties and/or theorems with this circularized random vector, its introduction is reasonable. Our approach for associating a circular random vector to a (possibly) noncircular one is motivated by the well-known method used for stationarizing a cyclostationary random process [33].

Definition 3.1: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the random vector $\mathbf{x}_{(\mathrm{a})} \stackrel{\triangle}{=} e^{j2\pi\psi}\mathbf{x}$, where $\psi \in [0, 1)$

is a uniformly distributed random variable independent of $\mathbf{x}$, is said to be *circular analog* of $\mathbf{x}$.

In the following, we will show that the circular analog is indeed a circular random vector. The next proposition expresses the distribution of $\mathbf{x}_{(\mathrm{a})}$ in terms of the distribution of $\mathbf{x}$ (for both polar and sheared-polar representations).

Proposition 3.2: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the pdfs of the polar representations and sheared-polar representations of $\mathbf{x}$ and its circular analog $\mathbf{x}_{(\mathrm{a})}$ are related according to

$$f_{\mathbf{x}^{(\mathrm{p})}_{(\mathrm{a})}}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n) = \int_0^1 f_{\mathbf{x}^{(\mathrm{p})}}\big(r_1,\ldots,r_n,[\phi_1 - \varphi]_{[0,1)},\ldots,[\phi_n-\varphi]_{[0,1)}\big)\,d\varphi \quad \lambda_{2n}\text{-a.e.} \tag{6a}$$

$$f_{\mathbf{x}^{(\mathrm{s})}_{(\mathrm{a})}}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n) = \int_0^1 f_{\mathbf{x}^{(\mathrm{s})}}\left(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n\right)d\phi_n \quad \lambda_{2n}\text{-a.e.} \tag{6b}$$

respectively.

Proof: For (6a), consider the joint pdf of $\mathbf{x}^{(\mathrm{p})}_{(\mathrm{a})}$ and ψ, i.e., $f_{\mathbf{x}^{(\mathrm{p})}_{(\mathrm{a})};\psi}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n,\varphi) = f_{\mathbf{x}^{(\mathrm{p})}_{(\mathrm{a})}|\psi}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n|\varphi)\,f_\psi(\varphi) = f_{\mathbf{x}^{(\mathrm{p})}}\big(r_1,\ldots,r_n,[\phi_1-\varphi]_{[0,1)},\ldots,[\phi_n-\varphi]_{[0,1)}\big)$ and marginalize with respect to φ. Equation (6b) follows from (6a) using Lemma 2.5 and identity (1). □

Observe that $f_{\mathbf{x}^{(\mathrm{s})}_{(\mathrm{a})}}$ does not depend on ϕ_n λ_{2n}-a.e., so that Corollary 2.6 implies circularity of $\mathbf{x}_{(\mathrm{a})}$.

A. Divergence Characterization

Here, we present a (nontrivial) characterization of the circular analog of a complex-valued random vector that further supports the chosen definition. It is based on the *Kullback–Leibler divergence* (or *relative entropy*) [34], [35], which can be regarded as a distance measure between two probability measures. For complex-valued random vectors, whose real representations are distributed according to multivariate pdfs, the Kullback–Leibler divergence $D(\mathbf{x}\|\mathbf{y}) \in \mathbb{R}_0^+ \cup \{\infty\}$ between $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^n$ is defined as

$$D\big(\mathbf{x}\|\mathbf{y}\big) \triangleq D\big(\mathbf{x}^{(\mathrm{r})}\|\mathbf{y}^{(\mathrm{r})}\big) \triangleq \int_{\mathbb{R}^{2n}} f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi})} d\boldsymbol{\xi}$$

where we set $0\log 0 \triangleq 0$ and $0 \log \frac{0}{0} \triangleq 0$ (motivated by continuity). Here, $D(\mathbf{x}\|\mathbf{y})$ is finite only if the support set of $f_{\mathbf{x}^{(\mathrm{r})}}$ is contained in the support set of $f_{\mathbf{y}^{(\mathrm{r})}}$ λ_{2n}-a.e.. Note that $D(\mathbf{x}\|\mathbf{y}) = 0$ if and only if $f_{\mathbf{x}^{(\mathrm{r})}} = f_{\mathbf{y}^{(\mathrm{r})}}$ λ_{2n}-a.e. [34]. The next lemma shows that $D(\mathbf{x}\|\mathbf{y})$ can be equivalently expressed in terms of polar and sheared-polar representations.

Lemma 3.3: Suppose $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^n$ are complex-valued random vectors. Then, the Kullback–Leibler divergence $D(\mathbf{x}\|\mathbf{y})$ can be computed from the respective polar and sheared-polar representations of $\mathbf{x}$ and $\mathbf{y}$ according to

$$\begin{aligned} D\big(\mathbf{x}\|\mathbf{y}\big) &= D\big(\mathbf{x}^{(\mathrm{p})}\|\mathbf{y}^{(\mathrm{p})}\big) = \int_{(\mathbb{R}_0^+)^n\times[0,1)^n} f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(\mathrm{p})}}(\boldsymbol{\xi})} d\boldsymbol{\xi} \\ &= D\big(\mathbf{x}^{(\mathrm{s})}\|\mathbf{y}^{(\mathrm{s})}\big) = \int_{(\mathbb{R}_0^+)^n\times[0,1)^n} f_{\mathbf{x}^{(\mathrm{s})}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(\mathrm{s})}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(\mathrm{s})}}(\boldsymbol{\xi})} d\boldsymbol{\xi}. \end{aligned}$$

Proof: With $\mathcal{A} \triangleq (\mathbb{R}_0^+)^n \times [0,1)^n$

$$\begin{aligned} D\big(\mathbf{x}^{(\mathrm{r})}\|\mathbf{y}^{(\mathrm{r})}\big) &= \int_{\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(\mathcal{A})} f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi})} d\boldsymbol{\xi} \\ &= \int_{\mathcal{A}} f_{\mathbf{x}^{(\mathrm{r})}}\Big(\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(\boldsymbol{\xi})\Big) \log \frac{f_{\mathbf{x}^{(\mathrm{r})}}\big(\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(\boldsymbol{\xi})\big)}{f_{\mathbf{y}^{(\mathrm{r})}}\big(\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}(\boldsymbol{\xi})\big)} \left|J_{\mathbb{T}^{(\mathrm{p}\to\mathrm{r})}}(\boldsymbol{\xi})\right| d\boldsymbol{\xi} \\ &= \int_{\mathcal{A}} f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(\mathrm{p})}}(\boldsymbol{\xi})} d\boldsymbol{\xi} = D\big(\mathbf{x}^{(\mathrm{p})}\|\mathbf{y}^{(\mathrm{p})}\big) \end{aligned}$$

where Lemma 2.3 has been used. Furthermore, using the mapping $\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}$ and the appropriate partition $\{\mathcal{A}_1,\ldots,\mathcal{A}_N\}$ of $\mathcal{A}$ (cf., the proof of Lemma 2.5)

$$\begin{aligned} D\big(\mathbf{x}^{(\mathrm{p})}\|\mathbf{y}^{(\mathrm{p})}\big) &= \int_{\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A})} f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(\mathrm{p})}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(\mathrm{p})}}(\boldsymbol{\xi})} d\boldsymbol{\xi} \\ &= \int_{\widetilde{\mathbb{T}}^{(\mathrm{p}\to\mathrm{s})}\left(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\mathcal{A})\right)} f_{\mathbf{x}^{(\mathrm{p})}}\Big(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\Big) \log \frac{f_{\mathbf{x}^{(\mathrm{p})}}\Big(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\Big)}{f_{\mathbf{y}^{(\mathrm{p})}}\Big(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\Big)} d\boldsymbol{\xi} \\ &= \bigcup_{i=1}^{N} \int_{\mathcal{A}_i} f_{\mathbf{x}^{(\mathrm{p})}}\Big(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi}+\mathbf{k}_i)\Big) \log \frac{f_{\mathbf{x}^{(\mathrm{p})}}\Big(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi}+\mathbf{k}_i)\Big)}{f_{\mathbf{y}^{(\mathrm{p})}}\Big(\widetilde{\mathbb{T}}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi}+\mathbf{k}_i)\Big)} d\boldsymbol{\xi} \\ &= \bigcup_{i=1}^{N} \int_{\mathcal{A}_i} f_{\mathbf{x}^{(\mathrm{p})}}\Big(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\Big) \log \frac{f_{\mathbf{x}^{(\mathrm{p})}}\big(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\big)}{f_{\mathbf{y}^{(\mathrm{p})}}\big(\mathbb{T}^{(\mathrm{s}\to\mathrm{p})}(\boldsymbol{\xi})\big)} d\boldsymbol{\xi} \\ &= D\big(\mathbf{x}^{(\mathrm{s})}\|\mathbf{y}^{(\mathrm{s})}\big) \end{aligned}$$

where Lemma 2.5 has been used. □

We intend to derive a theorem, which states that the circular analog has a smaller "distance" from the given complex-valued random vector than any other circular random vector. To that end, consider the sheared-polar representation of $\mathbf{x}$, i.e., $\mathbf{x}^{(\mathrm{s})} \in \mathbb{R}^{2n}$, and form the "reduced" vector $\tilde{\mathbf{x}}^{(\mathrm{s})} \in \mathbb{R}^{2n-1}$ by only taking the first $2n-1$ elements of $\mathbf{x}^{(\mathrm{s})}$. Clearly, its pdf is given by marginalization, i.e., $f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) = \int_0^1 f_{\mathbf{x}^{(\mathrm{s})}}(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n)d\phi_n$, where $\tilde{\boldsymbol{\xi}} \triangleq (r_1,\ldots,r_n,\phi_1,\ldots,\phi_{n-1})$. Furthermore, let $\widetilde{\mathcal{S}}_{\mathbf{x}} \subset (\mathbb{R}_0^+)^n \times [0,1)^{n-1}$ denote the support set of $f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}$.

Note that $f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) = 0$ is equivalent to $f_{\mathbf{x}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}, \phi_n) = 0$ λ_1-a.e. (for fixed $\tilde{\boldsymbol{\xi}}$). We have

$$f_{\mathbf{x}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}, \phi_n) = f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}), \quad \tilde{\boldsymbol{\xi}} \in \widetilde{\mathcal{S}}_{\mathbf{x}},\ \phi_n \in [0,1) \tag{7}$$

where $\vartheta \stackrel{\triangle}{=} \left(\mathbf{x}^{(\mathrm{s})}\right)_{2n}$ denotes the last element of $\mathbf{x}^{(\mathrm{s})}$.

Theorem 3.4: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, a circular random vector $\mathbf{y} \in \mathbb{C}^n$ is the circular analog of $\mathbf{x}$, i.e., $\mathbf{y} = \mathbf{x}_{(\mathrm{a})}$, if and only if it minimizes the Kullback–Leibler divergence to $\mathbf{x} \in \mathbb{C}^n$ within the whole set of circular random vectors, i.e., if and only if

$$D\big(\mathbf{x}\|\mathbf{y}\big) = \inf_{\mathbf{c}\in\mathcal{C}_n} D\big(\mathbf{x}\|\mathbf{c}\big).$$

Furthermore

$$\begin{aligned} D\big(\mathbf{x}\|\mathbf{x}_{(\mathrm{a})}\big) &= \inf_{\mathbf{c}\in\mathcal{C}_n} D\big(\mathbf{x}\|\mathbf{c}\big) \\ &= \int_{\widetilde{\mathcal{S}}_{\mathbf{x}}} f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) \left(\int_0^1 f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) \log f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) d\phi_n \right) d\tilde{\boldsymbol{\xi}} \\ &\stackrel{\triangle}{=} -h(\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}) \end{aligned}$$

where $h(\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})})$ denotes the *conditional differential entropy* of ϑ given $\tilde{\mathbf{x}}^{(\mathrm{s})}$, cf., [34] and Definition 4.2, with ϑ and $\tilde{\mathbf{x}}^{(\mathrm{s})}$ according to (7).

Proof: Suppose $\mathbf{c} \in \mathcal{C}_n$ and consider its sheared-polar representation $\mathbf{c}^{(\mathrm{s})} \in \mathbb{R}^{2n}$. Due to the circularity of $\mathbf{c}$, $f_{\mathbf{c}^{(\mathrm{s})}}(\boldsymbol{\xi}) = f_{\mathbf{c}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})$ λ_{2n}-a.e., and, according to Lemma 3.3

$$\begin{aligned} &D\big(\mathbf{x}\|\mathbf{c}\big) \\ &= \int_{\widetilde{\mathcal{S}}_{\mathbf{x}}\times[0,1)} f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})\, f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) \log \frac{f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})\, f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}})}{f_{\mathbf{c}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})} d\tilde{\boldsymbol{\xi}} d\phi_n \\ &\stackrel{(*)}{=} \int_{\widetilde{\mathcal{S}}_{\mathbf{x}}} f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) \log \frac{f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})}{f_{\mathbf{c}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})} d\tilde{\boldsymbol{\xi}} \\ &\quad + \int_{\widetilde{\mathcal{S}}_{\mathbf{x}}} f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) \left(\int_0^1 f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) \log f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) d\phi_n \right) d\tilde{\boldsymbol{\xi}} \\ &= D\big(\tilde{\mathbf{x}}^{(\mathrm{s})}\|\tilde{\mathbf{c}}^{(\mathrm{s})}\big) \\ &\quad + \int_{\widetilde{\mathcal{S}}_{\mathbf{x}}} f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) \left(\int_0^1 f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) \log f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) d\phi_n \right) d\tilde{\boldsymbol{\xi}} \end{aligned}$$

where $\tilde{\mathbf{c}}^{(\mathrm{s})} \in \mathbb{R}^{2n-1}$ is the corresponding "reduced" vector of $\mathbf{c}^{(\mathrm{s})}$. For the validity of $(*)$, we also refer to [35, Th. D.13]. It follows that

$$\begin{aligned} &\inf_{\mathbf{c}\in\mathcal{C}_n} D\big(\mathbf{x}\|\mathbf{c}\big) \\ &= \int_{\widetilde{\mathcal{S}}_{\mathbf{x}}} f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) \left(\int_0^1 f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) \log f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}(\phi_n|\tilde{\boldsymbol{\xi}}) d\phi_n \right) d\tilde{\boldsymbol{\xi}} \end{aligned}$$

and the infimum is achieved for $f_{\tilde{\mathbf{c}}^{(\mathrm{s})}} = f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}$ λ_{2n-1}-a.e.. Since for the circular analog $\mathbf{x}_{(\mathrm{a})}$ of $\mathbf{x}$, $f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}}(\boldsymbol{\xi}) = f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})$ λ_{2n}-a.e., and since $f_{\mathbf{c}^{(\mathrm{s})}}(\boldsymbol{\xi}) = f_{\tilde{\mathbf{c}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})$ λ_{2n}-a.e., the infimum is achieved if and only if $f_{\mathbf{c}^{(\mathrm{s})}} = f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}}$ λ_{2n}-a.e., i.e., $\mathbf{c} = \mathbf{x}_{(\mathrm{a})}$. □

B. Complex-Valued Random Vectors With Finite Second-Order Moments

In this section, we establish important properties of the circular analog $\mathbf{x}_{(\mathrm{a})}$ of a complex-valued random vector $\mathbf{x}$, whose second-order moments exist. Clearly, both mean vector and complementary covariance matrix of $\mathbf{x}_{(\mathrm{a})}$ are vanishing. For the covariance matrix, we have the following result.

Theorem 3.5: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a zero-mean complex-valued random vector with finite second-order moments. Then, the covariance matrix of the circular analog $\mathbf{x}_{(\mathrm{a})}$ equals the covariance matrix of $\mathbf{x}$, i.e., $\mathbf{C}_{\mathbf{x}_{(\mathrm{a})}} = \mathbf{C}_{\mathbf{x}}$.

Proof: For the correlation between the kth and lth entry of $\mathbf{x}_{(\mathrm{a})}$

$$\begin{aligned} &\mathrm{E}\left\{ \left(\mathbf{x}_{(\mathrm{a})}\right)_k \left(\mathbf{x}_{(\mathrm{a})}\right)_l^* \right\} \\ &= \int_{\mathbb{R}^{2n}} (\xi_k + j\xi_{k+n})(\xi_l - j\xi_{l+n}) f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{r})}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^{2n}} \int_0^1 (\xi_k + j\xi_{k+n})(\xi_l - j\xi_{l+n}) f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{r})}|\psi}(\boldsymbol{\xi}|\varphi) d\varphi d\boldsymbol{\xi} \\ &\stackrel{(*)}{=} \int_0^1 \int_{\mathbb{R}^{2n}} (\xi_k + j\xi_{k+n})(\xi_l - j\xi_{l+n}) f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{r})}|\psi}(\boldsymbol{\xi}|\varphi) d\boldsymbol{\xi} d\varphi \\ &= \int_0^1 \mathrm{E}\left\{ \left(e^{j2\pi\varphi}\mathbf{x}\right)_k \left(e^{j2\pi\varphi}\mathbf{x}\right)_l^* \right\} d\varphi \\ &= \mathrm{E}\left\{ (\mathbf{x})_k (\mathbf{x})_l^* \right\} \end{aligned}$$

where ψ denotes the uniformly distributed random variable used for the definition of $\mathbf{x}_{(\mathrm{a})}$ (see Definition 3.1) and $(*)$ follows from Fubini's Theorem [22]. □

The following theorem states that the circular analog of an improper Gaussian distributed random vector is non-Gaussian.

Theorem 3.6: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a zero-mean, complex-valued, and Gaussian distributed random vector with $\|\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}\|_2 < 1$ such that its circular analog $\mathbf{x}_{(\mathrm{a})}$ is Gaussian distributed. Here, $\mathbf{B}_{\mathbf{x}}$ denotes a generalized Cholesky factor of the nonsingular covariance matrix $\mathbf{C}_{\mathbf{x}}$ of $\mathbf{x}$ and $\mathbf{P}_{\mathbf{x}}$ denotes the complementary covariance matrix of $\mathbf{x}$. Then, $\mathbf{x}$ is proper.

Proof: We first prove the theorem for the special case $\mathbf{C}_{\mathbf{x}} = \mathbf{I}_n$ and $\mathbf{P}_{\mathbf{x}} = \boldsymbol{\Lambda}_{\mathbf{x}}$, where $\boldsymbol{\Lambda}_{\mathbf{x}} \in \mathbb{R}^{n\times n}$ denotes a diagonal matrix with nonnegative diagonal entries $\lambda_i < 1$. For fixed (deterministic) θ, consider the random vector $\mathbf{y}_{(\theta)} \stackrel{\triangle}{=} e^{j2\pi\theta}\mathbf{x}$, which has covariance matrix $\mathbf{C}_{\mathbf{y}_{(\theta)}} = \mathbf{I}_n$ and complementary covariance matrix $\mathbf{P}_{\mathbf{y}_{(\theta)}} = e^{j4\pi\theta}\boldsymbol{\Lambda}_{\mathbf{x}}$. According to (3), the covariance matrix of its real representation $\mathbf{y}_{(\theta)}^{(\mathrm{r})}$ is given by

$$\begin{aligned} \mathbf{C}_{\mathbf{y}_{(\theta)}^{(\mathrm{r})}} = \frac{1}{2} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} \\ + \frac{1}{2} \begin{bmatrix} \cos(4\pi\theta)\, \boldsymbol{\Lambda}_{\mathbf{x}} & \sin(4\pi\theta)\, \boldsymbol{\Lambda}_{\mathbf{x}} \\ \sin(4\pi\theta)\, \boldsymbol{\Lambda}_{\mathbf{x}} & -\cos(4\pi\theta) \boldsymbol{\Lambda}_{\mathbf{x}} \end{bmatrix} \end{aligned}$$

whose determinant is easily computed as

$$\det \mathbf{C}_{\mathbf{y}^{(\mathrm{r})}_{(\theta)}} = 2^{-2n} \prod_{i=1}^{n} (1-\lambda_i^2).$$

Furthermore, its inverse is calculated as

$$\mathbf{C}^{-1}_{\mathbf{y}^{(\mathrm{r})}_{(\theta)}} = 2 \begin{bmatrix} \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1} \end{bmatrix} - 2 \begin{bmatrix} \cos(4\pi\theta)\mathbf{D}_{\mathbf{x}} & \sin(4\pi\theta)\mathbf{D}_{\mathbf{x}} \\ \sin(4\pi\theta)\mathbf{D}_{\mathbf{x}} & -\cos(4\pi\theta)\mathbf{D}_{\mathbf{x}} \end{bmatrix}$$

where $\mathbf{D}_{\mathbf{x}} \triangleq \boldsymbol{\Lambda}_{\mathbf{x}} \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1}$. Therefore, the pdf of $\mathbf{y}^{(\mathrm{r})}_{(\theta)}$ is given by

$$f_{\mathbf{y}^{(\mathrm{r})}_{(\theta)}}(\boldsymbol{\xi}) = \frac{1}{\pi^n \prod_{i=1}^{n} \sqrt{1-\lambda_i^2}} \times \exp\left(-\boldsymbol{\xi}^T \begin{bmatrix} \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1} \end{bmatrix} \boldsymbol{\xi} \right) \times \exp\left(\boldsymbol{\xi}^T \begin{bmatrix} \cos(4\pi\theta)\,\mathbf{D}_{\mathbf{x}} & \sin(4\pi\theta)\,\mathbf{D}_{\mathbf{x}} \\ \sin(4\pi\theta)\,\mathbf{D}_{\mathbf{x}} & -\cos(4\pi\theta)\,\mathbf{D}_{\mathbf{x}} \end{bmatrix} \boldsymbol{\xi} \right)$$

λ_{2n}-a.e.. Since $f_{\mathbf{x}^{(\mathrm{r})}_{(\mathrm{a})}}(\boldsymbol{\xi}) = \int_0^1 f_{\mathbf{y}^{(\mathrm{r})}_{(\theta)}}(\boldsymbol{\xi}) d\theta$ λ_{2n}-a.e., we obtain

$$f_{\mathbf{x}^{(\mathrm{r})}_{(\mathrm{a})}}(\boldsymbol{\xi}) = \frac{1}{\pi^n \prod_{i=1}^{n} \sqrt{1-\lambda_i^2}} \times \exp\left(-\boldsymbol{\xi}^T \begin{bmatrix} \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I}_n - \boldsymbol{\Lambda}_{\mathbf{x}}^2\right)^{-1} \end{bmatrix} \boldsymbol{\xi} \right) \times I_0\left(\left(\boldsymbol{\xi}^T \begin{bmatrix} \mathbf{D}_{\mathbf{x}} & \mathbf{0} \\ \mathbf{0} & -\mathbf{D}_{\mathbf{x}} \end{bmatrix} \boldsymbol{\xi} \right)^2 + \left(\boldsymbol{\xi}^T \begin{bmatrix} \mathbf{0} & \mathbf{D}_{\mathbf{x}} \\ \mathbf{D}_{\mathbf{x}} & \mathbf{0} \end{bmatrix} \boldsymbol{\xi} \right)^2 \right) \quad \lambda_{2n}\text{-a.e.} \tag{8}$$

where $I_0(x) = \int_0^1 \exp\left(x \cos(2\pi\theta)\right) d\theta$ is the modified Bessel function of the first kind of order zero [36]. Here, we have used the identity

$$\begin{aligned} &\int_0^1 \exp\left(a\cos(2\pi\theta) + b\sin(2\pi\theta)\right) d\theta \\ &= \int_0^1 \exp\left(r\cos(2\pi\theta_0)\cos(2\pi\theta) + r\sin(2\pi\theta_0)\sin(2\pi\theta)\right) d\theta \\ &= \int_0^1 \exp\left(r\cos(2\pi(\theta-\theta_0))\right) d\theta \\ &= I_0\left(\sqrt{a^2+b^2}\right) \end{aligned}$$

where (r, θ_0) denotes the polar coordinates of the complex number $(a + jb)$. According to the assumptions of the theorem and to Theorem 3.5, $\mathbf{x}_{(\mathrm{a})}$ is Gaussian distributed with covariance matrix $\mathbf{C}_{\mathbf{x}_{(\mathrm{a})}} = \mathbf{I}_n$ and vanishing mean vector and complementary covariance matrix. Hence, (8) implies $\mathbf{D}_{\mathbf{x}} = 0$ and, therefore, the statement.

For the general case, apply the Takagi factorization (5) to $\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}$, i.e., $\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T} = \mathbf{Q}_{\mathbf{x}}\boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{Q}_{\mathbf{x}}^T$, and consider the random vector $\mathbf{y} \triangleq \mathbf{Q}_{\mathbf{x}}^{-1}\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{x}$. Clearly, $\mathbf{C}_{\mathbf{y}} = \mathbf{I}_n$ and $\mathbf{P}_{\mathbf{y}} = \boldsymbol{\Lambda}_{\mathbf{x}}$, and both $\mathbf{y}$ and $\mathbf{y}_{(\mathrm{a})}$ are Gaussian distributed. From the special case, $\mathbf{P}_{\mathbf{y}} = \mathbf{0}$, and, in turn, $\mathbf{P}_{\mathbf{x}} = \mathbf{B}_{\mathbf{x}}\mathbf{Q}_{\mathbf{x}}\mathbf{P}_{\mathbf{y}}\mathbf{Q}_{\mathbf{x}}^T\mathbf{B}_{\mathbf{x}}^T = \mathbf{0}$. □

IV. Differential Entropy of Complex-Valued Random Vectors

As outlined in Section I, we are interested in bounds on the differential entropy of complex-valued random vectors. We start with a series of definitions, which are required for the further development of the paper. Again, we make use of the convention $0 \log 0 \triangleq 0$ and $0 \log \frac{0}{0} \triangleq 0$.

Definition 4.1: The *differential entropy* $h(\mathbf{x})$ of a complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is defined as the differential entropy of its real representation $\mathbf{x}^{(\mathrm{r})}$, i.e.,

$$h(\mathbf{x}) \triangleq h\left(\mathbf{x}^{(\mathrm{r})}\right) \triangleq -\int_{\mathbb{R}^{2n}} f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) \log f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

provided that the integrand is integrable [20].

Definition 4.2: The *conditional differential entropy* $h(\mathbf{x}|\mathbf{y})$ of a complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ given a complex valued random vector $\mathbf{y} \in \mathbb{C}^m$ is defined as the conditional differential entropy of the real representation $\mathbf{x}^{(\mathrm{r})}$ given the real representation $\mathbf{y}^{(\mathrm{r})}$, i.e.,

$$h(\mathbf{x}|\mathbf{y}) \triangleq h\left(\mathbf{x}^{(\mathrm{r})}|\mathbf{y}^{(\mathrm{r})}\right) \triangleq -\int_{\mathbb{R}^{2n+2m}} f_{\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi},\boldsymbol{\eta}) \log \frac{f_{\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi},\boldsymbol{\eta})}{f_{\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\eta})} d\boldsymbol{\xi} d\boldsymbol{\eta}$$

provided that the integrand is integrable. Here, $f_{\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi},\boldsymbol{\eta})$ denotes the joint pdf of $\mathbf{x}^{(\mathrm{r})}$ and $\mathbf{y}^{(\mathrm{r})}$, whereas $f_{\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\eta})$ denotes the marginal pdf of $\mathbf{y}^{(\mathrm{r})}$.

Definition 4.3: The *mutual information* $I(\mathbf{x};\mathbf{y})$ between the complex-valued random vectors $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$ is defined as the mutual information between their real representations $\mathbf{x}^{(\mathrm{r})}$ and $\mathbf{y}^{(\mathrm{r})}$, i.e.,

$$I(\mathbf{x};\mathbf{y}) \triangleq I\left(\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}\right) \triangleq \int_{\mathbb{R}^{2n+2m}} f_{\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi},\boldsymbol{\eta}) \log \frac{f_{\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi},\boldsymbol{\eta})}{f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) f_{\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\eta})} d\boldsymbol{\xi} d\boldsymbol{\eta}$$

where $f_{\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\xi},\boldsymbol{\eta})$ denotes the joint pdf of $\mathbf{x}^{(\mathrm{r})}$ and $\mathbf{y}^{(\mathrm{r})}$, and $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})$ and $f_{\mathbf{y}^{(\mathrm{r})}}(\boldsymbol{\eta})$ are the marginal pdfs of $\mathbf{x}^{(\mathrm{r})}$ and $\mathbf{y}^{(\mathrm{r})}$, respectively. It is well known that these quantities satisfy the following relations:

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \tag{9a}$$

$$I(\mathbf{x};\mathbf{y}) \geq 0 \tag{9b}$$

with equality in (9b) if and only if $\mathbf{x}$ and $\mathbf{y}$ are statistically independent. Furthermore, according to the next theorem, Gaussian distributed circular/proper random vectors are known to be entropy maximizers.

Theorem 4.4 [14]: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a zero-mean complex-valued random vector with nonsingular covariance matrix $\mathbf{C_x}$. Then, the differential entropy of $\mathbf{x}$ satisfies

$$h(\mathbf{x}) \leq \log\det(\pi e \mathbf{C_x}) \tag{10}$$

with equality if and only if $\mathbf{x}$ is Gaussian distributed and circular/proper.

Proof: See, e.g., [14] or [15]. □

Remarks: Let us assume, for the moment, that $\mathbf{x}$ is known to be non-Gaussian. Clearly, inequality (10) is strict in this case and $\log\det(\pi e \mathbf{C_x})$ is not a tight upper bound for the differential entropy $h(\mathbf{x})$. Similarly, if $\mathbf{x}$ is known to be improper, the differential entropy $h(\mathbf{x})$ is strictly smaller than $\log\det(\pi e \mathbf{C_x})$. Loosely speaking, there are two sources that decrease the differential entropy of a complex-valued random vector: non-Gaussianity and improperness. In the following, we will derive improved maximum entropy theorems that take this observation into account. While their application is not limited to the non-Gaussian and improper cases, the obtained upper bounds are in general tighter for these two scenarios than the upper bound given by Theorem 4.4.

A. Maximum Entropy Theorem I

We first prove a maximum entropy theorem that is especially suited to the non-Gaussian case. However, also for Gaussian distributed random vectors, the obtained upper bound will turn out to be tighter than the one of Theorem 4.4. It associates a specific circular random vector to a given random vector and upper bounds the differential entropy of the given random vector by the differential entropy of the associated circular random vector.

Theorem 4.5 (Maximum Entropy Theorem I): Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the differential entropies of $\mathbf{x}$ and its circular analog $\mathbf{x}_{(\mathrm{a})}$ satisfy

$$h(\mathbf{x}) \leq h\left(\mathbf{x}_{(\mathrm{a})}\right)$$

with equality if and only if $\mathbf{x}$ is circular.

Proof: Since $\mathbf{x}_{(\mathrm{a})} = e^{j2\pi\psi}\mathbf{x}$ with ψ independent of $\mathbf{x}$, we have for fixed (deterministic) φ

$$h(\mathbf{x}_{(\mathrm{a})}|\psi = \varphi) = h(e^{j2\pi\varphi}\mathbf{x}) = h(\mathbf{x})$$

and, furthermore, by applying Fubini's theorem to Definition 4.2

$$\begin{aligned} h(\mathbf{x}_{(\mathrm{a})}|\psi) &= \int h(\mathbf{x}_{(\mathrm{a})}|\psi = \varphi) f_\psi(\varphi) d\varphi \\ &= \int h(\mathbf{x}) f_\psi(\varphi) d\varphi = h(\mathbf{x}). \end{aligned}$$

Therefore

$$h(\mathbf{x}_{(\mathrm{a})}) - h(\mathbf{x}) = h(\mathbf{x}_{(\mathrm{a})}) - h(\mathbf{x}_{(\mathrm{a})}|\psi) = I(\mathbf{x}_{(\mathrm{a})};\psi) \geq 0 \tag{11}$$

where we have used (9a) and (9b). $\mathbf{x}_{(\mathrm{a})}$ and ψ are independent, i.e., $h(\mathbf{x}_{(\mathrm{a})}) = h(\mathbf{x})$, if and only if $\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}$ and ψ are independent. To investigate this independence,[3] consider the joint pdf of $\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}$ and ψ, i.e.,

$$\begin{aligned} f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})};\psi}&(r_1,\ldots,r_n,\phi_1,\ldots,\phi_n,\varphi) \\ &= f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}|\psi}(r_1,\ldots\ldots,r_n,\phi_1,\ldots,\phi_n|\varphi)\, f_\psi(\varphi) \\ &= f_{\mathbf{x}^{(\mathrm{s})}}\left(r_1,\ldots,r_n,\phi_1,\ldots,\phi_{n-1},[\phi_n-\varphi]_{[0,1)}\right) \\ &= f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}}) f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}\left([\phi_n-\varphi]_{[0,1)}\middle|\tilde{\boldsymbol{\xi}}\right) \end{aligned}$$

where $\tilde{\boldsymbol{\xi}} \triangleq (r_1,\ldots,r_n,\phi_1,\ldots,\phi_{n-1}) \in \widetilde{\mathcal{S}}_\mathbf{x}$, $\widetilde{\mathcal{S}}_\mathbf{x} \subset (\mathbb{R}_0^+)^n \times [0,1)^{n-1}$ being the support set of $f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}$, cf., (7). Since $f_{\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}},\phi_n) = f_{\tilde{\mathbf{x}}^{(\mathrm{s})}}(\tilde{\boldsymbol{\xi}})$ λ_{2n}-a.e. on $\widetilde{\mathcal{S}}_\mathbf{x} \times [0,1)$ and $f_\psi(\varphi) = 1$ λ_1-a.e. on $[0,1)$, independence of $\mathbf{x}_{(\mathrm{a})}^{(\mathrm{s})}$ and ψ is equivalent to

$$f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}\left([\phi_n-\varphi]_{[0,1)}\middle|\tilde{\boldsymbol{\xi}}\right) = 1 \quad \lambda_{2n+1}\text{-a.e. on } [0,1)^2 \times \widetilde{\mathcal{S}}_\mathbf{x} \tag{12}$$

as function of $(\phi_n,\varphi,\tilde{\boldsymbol{\xi}})$. Transforming both sides of this equation according to $\phi'_n \triangleq [\phi_n-\varphi]_{[0,1)}$ and $\varphi' \triangleq \varphi$, a similar partitioning argument as in the proof of Lemma 2.5 shows that (12) is equivalent to

$$f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}\left(\phi'_n\middle|\tilde{\boldsymbol{\xi}}\right) = 1 \quad \lambda_{2n+1}\text{-a.e. on } [0,1)^2 \times \widetilde{\mathcal{S}}_\mathbf{x} \tag{13}$$

as function of $(\phi'_n,\varphi',\tilde{\boldsymbol{\xi}})$. Marginalization of both sides of (13) with respect to φ' yields

$$f_{\vartheta|\tilde{\mathbf{x}}^{(\mathrm{s})}}\left(\phi'_n\middle|\tilde{\boldsymbol{\xi}}\right) = 1 \quad \lambda_{2n}\text{-a.e. on } [0,1) \times \widetilde{\mathcal{S}}_\mathbf{x}$$

as function of $(\phi'_n,\tilde{\boldsymbol{\xi}})$, so that—according to Corollary 2.6 and (7)—equality $h(\mathbf{x}_{(\mathrm{a})}) = h(\mathbf{x})$ implies circularity of $\mathbf{x}$. The converse statement follows from Theorem 3.4. □

Remarks: Since $\mathbf{x}_{(\mathrm{a})}$ is non-Gaussian in general,[4] the upper bound in Theorem 4.5 is typically tighter than the upper bound in Theorem 4.4. Furthermore, Theorem 4.5 does not need the requirement of finite second-order moments. The next corollary states that for improper Gaussian distributed random vectors the upper bound in Theorem 4.5 is strictly smaller than the upper bound in Theorem 4.4.

Corollary 4.6: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a zero-mean, complex-valued, and Gaussian distributed random vector with nonsingular covariance matrix $\mathbf{C_x}$ and $\|\mathbf{B_x}^{-1}\mathbf{P_x}\mathbf{B_x}^{-T}\|_2 < 1$, where $\mathbf{B_x}$ denotes a generalized Cholesky factor of $\mathbf{C_x}$ and $\mathbf{P_x}$ denotes the complementary covariance matrix of $\mathbf{x}$. Then, the differential entropy of its circular analog $\mathbf{x}_{(\mathrm{a})}$ satisfies

$$h\left(\mathbf{x}_{(\mathrm{a})}\right) \leq \log\det(\pi e \mathbf{C_x})$$

with equality if and only if $\mathbf{x}$ is proper.

Proof: Since $\mathbf{x}_{(\mathrm{a})}$ is zero-mean with covariance matrix $\mathbf{C}_{\mathbf{x}_{(\mathrm{a})}} = \mathbf{C_x}$, cf., Theorem 3.5, the inequality follows from

[3] The following technical derivation is required in order to show identical distributions of $e^{j2\pi\theta}\mathbf{x}$ for all $\theta \in [0,1)$ and not only for θ λ_1-a.e. on $[0,1)$.

[4] Note that it is possible to define an improper (non-Gaussian, cf., Theorem 3.6) random vector, such that its circular analog is Gaussian distributed. In this case, Theorem 4.5 does not yield an improvement over Theorem 4.4.

Theorem 4.4. Furthermore, equality $h\left(\mathbf{x}_{(\mathrm{a})}\right) = \log\det(\pi e \mathbf{C}_{\mathbf{x}})$ implies Gaussianity of $\mathbf{x}_{(\mathrm{a})}$, and, according to Theorem 3.6, properness of $\mathbf{x}$. □

B. Maximum Entropy Theorem II

Here, we prove a maximum entropy theorem that is especially suited to the improper case. The derivation is based on a maximum entropy theorem for *real-valued* random vectors.

Theorem 4.7: Suppose $\mathbf{x} \in \mathbb{R}^n$ is a *real-valued* random vector with nonsingular covariance matrix $\mathbf{C}_{\mathbf{x}}$. Then, the differential entropy of $\mathbf{x}$ satisfies

$$h(\mathbf{x}) \le \frac{1}{2}\log\det\left(2\pi e \mathbf{C}_{\mathbf{x}}\right)$$

with equality if and only if $\mathbf{x}$ is Gaussian distributed.

Proof: For the proof of this theorem for $\mathbf{x}$ being zero-mean see e.g., [34]. The general case, where $\mathbf{x}$ has a non-vanishing mean vector, follows immediately since both differential entropy and covariance matrix are invariant with respect to translations. □

We are now able to state the main theorem of Section IV-B.

Theorem 4.8 (Maximum Entropy Theorem II): Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector with nonsingular covariance matrix $\mathbf{C}_{\mathbf{x}}$ and $\|\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}\|_2 < 1$, where $\mathbf{B}_{\mathbf{x}}$ denotes a generalized Cholesky factor of $\mathbf{C}_{\mathbf{x}}$ and $\mathbf{P}_{\mathbf{x}}$ denotes the complementary covariance matrix of $\mathbf{x}$. Then, the differential entropy of $\mathbf{x}$ satisfies

$$h(\mathbf{x}) \le \log\det(\pi e \mathbf{C}_{\mathbf{x}}) + \frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2)$$

where λ_i are the singular values of $\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}$, with equality if and only if $\mathbf{x}$ is Gaussian distributed.

Proof: According to Theorem 4.7

$$\begin{aligned}
h(\mathbf{x}) &\le \frac{1}{2}\log\det\left(2\pi e \mathbf{C}_{\mathbf{x}^{(\mathrm{r})}}\right)\\
&\overset{(3)}{=} \frac{1}{2}\log\det\left(\pi e\left(\overline{\mathbf{C}}_{\mathbf{x}} + \underline{\mathbf{P}}_{\mathbf{x}}\right)\right)\\
&= n\log(\pi e) + \frac{1}{2}\log\det\left(\overline{\mathbf{B}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{H}} + \underline{\mathbf{P}}_{\mathbf{x}}\right)\\
&\overset{(4a),(4c)}{=} n\log(\pi e) + \frac{1}{2}\log\det\left(\overline{\mathbf{B}}_{\mathbf{x}}\overline{\mathbf{B}}_{\mathbf{x}}^{T} + \underline{\mathbf{P}}_{\mathbf{x}}\right)\\
&\overset{(4e)}{=} \log\det(\pi e \mathbf{C}_{\mathbf{x}})\\
&\qquad + \frac{1}{2}\log\det\left(\mathbf{I}_{2n} + \overline{\mathbf{B}}_{\mathbf{x}}^{-1}\underline{\mathbf{P}}_{\mathbf{x}}\overline{\mathbf{B}}_{\mathbf{x}}^{-T}\right)\\
&\overset{(4a),(4b),(4c)}{=} \log\det(\pi e \mathbf{C}_{\mathbf{x}})\\
&\qquad + \frac{1}{2}\log\det\left(\mathbf{I}_{2n} + \underline{\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}}\right)\\
&= \log\det(\pi e \mathbf{C}_{\mathbf{x}}) + \frac{1}{2}\log\prod_{i=1}^{n}(1-\lambda_i^2)
\end{aligned}$$

where the last identity follows from Proposition 2.8 applied to the random vector $\mathbf{y} \overset{\triangle}{=} \mathbf{B}_{\mathbf{x}}^{-1}\mathbf{x}$. Note that $\mathbf{P}_{\mathbf{y}} = \mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}$. We also conclude from the last expression that the nonsingularity of $\mathbf{C}_{\mathbf{x}^{(\mathrm{r})}}$, which is required for the application of Theorem 4.7, is a direct consequence of the assumption $\|\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}\|_2 < 1$, since $\lambda_i \le \|\mathbf{B}_{\mathbf{x}}^{-1}\mathbf{P}_{\mathbf{x}}\mathbf{B}_{\mathbf{x}}^{-T}\|_2$. The equality criterion is obvious. □

Remarks: Note that $\frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2) \le 0$ with equality if and only if $\mathbf{x}$ is proper, so that Theorem 4.8 implies Theorem 4.4. The upper bound in Theorem 4.8 is the differential entropy of a Gaussian distributed but in general noncircular/improper random vector with same covariance matrix and complementary covariance matrix as $\mathbf{x}$, whereas the upper bound in Theorem 4.5 is the differential entropy of a circular but in general non-Gaussian random vector with same covariance matrix as $\mathbf{x}$. Which of the two bounds is tighter depends on the situation, i.e., on the degree of improperness and non-Gaussianity; a general statement is not possible. However, for an improper Gaussian distributed random vector $\mathbf{x}$

$$\log\det(\pi e \mathbf{C}_{\mathbf{x}}) + \frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2) < h(\mathbf{x}_{(\mathrm{a})})$$

whereas for a circular non-Gaussian random vector $\mathbf{x}$

$$\begin{aligned}
h(\mathbf{x}_{(\mathrm{a})}) &< \log\det(\pi e \mathbf{C}_{\mathbf{x}}) + \frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2)\\
&= \log\det(\pi e \mathbf{C}_{\mathbf{x}}).
\end{aligned}$$

V. Capacity of Complex-Valued Channels

In this section we study the influence of circularity/properness—noncircularity/improperness on channel capacity. In particular, we investigate MIMO channels with complex-valued input and complex-valued output. For simplicity, we only consider linear channels with additive noise, i.e., channels of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{14}$$

where $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{y} \in \mathbb{C}^n$, and $\mathbf{z} \in \mathbb{C}^n$ denote transmit, receive, and noise vector, respectively, and $\mathbf{H} \in \mathbb{C}^{n\times m}$ is the channel matrix. Both[5] $\mathbf{x}$ and $\mathbf{z}$ are modeled as independent identically distributed (i.i.d.) (only with respect to channel uses; within the random vectors the i.i.d. assumption is not made) vector-valued random processes, whereas $\mathbf{H}$ is either assumed to be deterministic or is modeled as an i.i.d. (again, only with respect to channel uses) matrix-valued random process. Furthermore, $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{H}$ (if applicable) are assumed to be statistically independent. Note that the assumption of a Gaussian distributed noise vector $\mathbf{z}$ is only made for the special case investigated in Section V-C but not in general. The channel is characterized by the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ via the conditional pdf $f_{\mathbf{y}^{(\mathrm{r})}|\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\eta}|\boldsymbol{\xi})$ of their real representations $\mathbf{y}^{(\mathrm{r})}$ given $\mathbf{x}^{(\mathrm{r})}$, as well as by a set $\mathcal{I}$ of admissible input distributions. We write $\mathbf{x} \in \mathcal{I}$, if the distribution of $\mathbf{x}$ defined by the pdf $f_{\mathbf{x}^{(\mathrm{r})}}$ is in $\mathcal{I}$. Then, the *capacity/noncoherent capacity* of (14) is given by the supremum of the mutual information over the set of admissible input distributions [37], i.e., by

$$C = \sup_{\mathbf{x}\in\mathcal{I}} I(\mathbf{x};\mathbf{y}).$$

[5]Without loss of generality, since an i.i.d. (with respect to channel uses) $\mathbf{x}$ is capacity achieving if $\mathbf{z}$ and $\mathbf{H}$ (if applicable) are i.i.d..

If, for the case of a random channel matrix, it is additionally assumed that the channel realizations are known to the receiver (but not to the transmitter), the channel output of (14) is the pair

$$(\mathbf{y},\mathbf{H}) = (\mathbf{H}\mathbf{x}+\mathbf{z},\mathbf{H}) \tag{15}$$

so that the channel law of (15) is governed by the conditional pdf $f_{\mathbf{y}^{(\mathrm{r})};\mathbf{H}^{(\mathrm{r})}|\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\eta},\boldsymbol{\chi}|\boldsymbol{\xi})$, where $\mathbf{H}^{(\mathrm{r})}$ is defined by an appropriate stacking of real and imaginary part of $\mathbf{H}$. Therefore, the *coherent capacity* of (15) is given by

$$\begin{aligned} C_{\mathrm{c}} &= \sup_{\mathbf{x}\in\mathcal{I}} I(\mathbf{x};\mathbf{y},\mathbf{H}) \\ &= \sup_{\mathbf{x}\in\mathcal{I}} \int I(\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}\big|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \end{aligned}$$

where $f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})$ denotes the pdf of $\mathbf{H}^{(\mathrm{r})}$ and Fubini's Theorem has been used. A random vector $\mathbf{x}\in\mathcal{I}$ is said to be *capacity-achieving* for (14) or (15), if $I(\mathbf{x};\mathbf{y}) = C$ or $I(\mathbf{x};\mathbf{y},\mathbf{H}) = C_{\mathrm{c}}$, respectively. In what follows (see Sections V-A and V-B), the existence of a (not necessarily circular) capacity-achieving random vector $\mathbf{x}\in\mathcal{I}$ is presupposed.

A. Circular Noise Vector

Here, we assume that the noise vector $\mathbf{z}\in\mathbb{C}^n$ is circular and that $\mathcal{I}$ is closed under the operation of forming the circular analog, i.e., that $\mathbf{x}\in\mathcal{I}$ implies $\mathbf{x}_{(\mathrm{a})}\in\mathcal{I}$—in the following shortly termed *circular-closed*. Note that this closeness assumption is a natural assumption, since the operation of forming the circular analog preserves both peek and average power constraints, cf., Theorem 3.5, which are the most common constraints for defining $\mathcal{I}$. If $\mathbf{z}$ is Gaussian distributed, it has been shown in [15] that capacity (for deterministic $\mathbf{H}$) and coherent capacity (for random $\mathbf{H}$) are achieved by circular (Gaussian distributed) random vectors, respectively. The proofs are based on Theorem 4.4. The following theorems extend these results to the non-Gaussian case.

Theorem 5.1: Suppose for (14) a deterministic channel matrix $\mathbf{H}\in\mathbb{C}^{n\times m}$, a circular noise vector $\mathbf{z}\in\mathbb{C}^n$, and a circular-closed set $\mathcal{I}$ of admissible input distributions. Then, there exists a circular random vector $\mathbf{x}\in\mathbb{C}^m$ that achieves the capacity of (14).

Proof: Let us denote by $\mathbf{x}'\in\mathcal{I}$ a—not necessarily circular—capacity-achieving random vector. According to (9a), its circular analog $\mathbf{x}\stackrel{\triangle}{=}\mathbf{x}'_{(\mathrm{a})} = e^{j2\pi\psi}\mathbf{x}'\in\mathcal{I}$, where $\psi\in[0,1)$ is uniformly distributed and assumed to be independent of $\mathbf{x}'$ and $\mathbf{z}$, satisfies

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \\ &= h(\mathbf{H}\mathbf{x}+\mathbf{z}) - h(\mathbf{H}\mathbf{x}+\mathbf{z}|\mathbf{x}) \\ &= h\left(\mathbf{H}e^{j2\pi\psi}\mathbf{x}'+\mathbf{z}\right) - h(\mathbf{z}) \\ &= h\left(e^{j2\pi\psi}(\mathbf{H}\mathbf{x}'+e^{-j2\pi\psi}\mathbf{z})\right) - h(\mathbf{z}). \end{aligned}$$

Note that $\mathbf{z}_{(\mathrm{a})} = e^{-j2\pi\psi}\mathbf{z} = \mathbf{z}$, cf., Theorem 3.4, and that $\mathbf{z}_{(\mathrm{a})}$ is independent of ψ, according to (11). Therefore

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= h\left((\mathbf{H}\mathbf{x}'+\mathbf{z})_{(\mathrm{a})}\right) - h(\mathbf{z}) \\ &\stackrel{(*)}{\geq} h(\mathbf{H}\mathbf{x}'+\mathbf{z}) - h(\mathbf{z}) \\ &= I(\mathbf{x}';\mathbf{H}\mathbf{x}'+\mathbf{z}) \\ &= C \end{aligned}$$

where $(*)$ follows from Theorem 4.5. Hence, the circular $\mathbf{x}$ is capacity-achieving. □

Theorem 5.2: Suppose for (14) a random channel matrix $\mathbf{H}\in\mathbb{C}^{n\times m}$, a circular noise vector $\mathbf{z}\in\mathbb{C}^n$, and a circular-closed set $\mathcal{I}$ of admissible input distributions. Then, there exists a circular random vector $\mathbf{x}\in\mathbb{C}^m$ that achieves the noncoherent capacity of (14).

Proof: Let us denote by $\mathbf{x}'\in\mathcal{I}$ a—not necessarily circular—capacity-achieving random vector and let $\mathbf{x}_{(\theta)}\stackrel{\triangle}{=}e^{j2\pi\theta}\mathbf{x}'$ (with $\theta\in[0,1)$ being deterministic). With $\mathbf{y}_{(\theta)} = \mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}$, we obtain

$$\begin{aligned} &I\left(\mathbf{x}_{(\theta)};\mathbf{y}_{(\theta)}\right) \\ &= h\left(\mathbf{y}_{(\theta)}\right) - h\left(\mathbf{y}_{(\theta)}\big|\mathbf{x}_{(\theta)}\right) \\ &= h\left(\mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}\right) - h\left(\mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}\big|\mathbf{x}_{(\theta)}\right) \\ &\stackrel{(*)}{=} h\left(e^{j2\pi\theta}(\mathbf{H}\mathbf{x}'+e^{-j2\pi\theta}\mathbf{z})\right) \\ &\quad - \int h\left(\mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}\big|\mathbf{x}_{(\theta)}^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}_{(\theta)}^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= h\left(e^{j2\pi\theta}(\mathbf{H}\mathbf{x}'+e^{-j2\pi\theta}\mathbf{z})\right) \\ &\quad - \int h\left(\mathbf{H}e^{j2\pi\theta}\mathbf{x}'+\mathbf{z}\big|\mathbf{x}'^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}'^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= h\left(e^{j2\pi\theta}(\mathbf{H}\mathbf{x}'+e^{-j2\pi\theta}\mathbf{z})\right) \\ &\quad - \int h\left(e^{j2\pi\theta}(\mathbf{H}\mathbf{x}'+e^{-j2\pi\theta}\mathbf{z})\big|\mathbf{x}'^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}'^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \end{aligned}$$

where $(*)$ follows from Fubini's Theorem. Since the differential entropy of a complex-valued random vector is invariant with respect to a multiplication with $e^{j2\pi\theta}$

$$\begin{aligned} &I\left(\mathbf{x}_{(\theta)};\mathbf{y}_{(\theta)}\right) \\ &= h\left(\mathbf{H}\mathbf{x}'+e^{-j2\pi\theta}\mathbf{z}\right) \\ &\quad - \int h\left(\mathbf{H}\mathbf{x}'+e^{-j2\pi\theta}\mathbf{z}\big|\mathbf{x}'^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}'^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &\stackrel{(*)}{=} h(\mathbf{H}\mathbf{x}'+\mathbf{z}) - \int h\left(\mathbf{H}\mathbf{x}'+\mathbf{z}\big|\mathbf{x}'^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}'^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= I(\mathbf{x}';\mathbf{y}') \end{aligned}$$

where $\mathbf{y}' = \mathbf{H}\mathbf{x}'+\mathbf{z}$ and $(*)$ follows from the circularity of $\mathbf{z}$. Hence, $\mathbf{x}_{(\theta)}$ is capacity-achieving. It is well known that the mutual information is a concave function with respect to the input distribution for fixed channel law [37]. Therefore, by Jensen's inequality [38], the random vector $\mathbf{x}\in\mathbb{C}^m$ with distribution defined according to $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})\stackrel{\triangle}{=}\int_0^1 f_{\mathbf{x}_{(\theta)}^{(\mathrm{r})}}(\boldsymbol{\xi})d\theta\ \lambda_{2n}$-a.e. satisfies

$$I(\mathbf{x};\mathbf{y}) \geq \int_0^1 I\left(\mathbf{x}_{(\theta)};\mathbf{y}_{(\theta)}\right)d\theta = \int_0^1 I(\mathbf{x}';\mathbf{y}')\,d\theta = I(\mathbf{x}';\mathbf{y}')$$

so that $\mathbf{x}$ achieves the noncoherent capacity of (14). But $f_{\mathbf{x}_{(\theta)}^{(\mathrm{r})}}(\boldsymbol{\xi}) = f_{\mathbf{x}'^{(\mathrm{r})}_{(\mathrm{a})}|\psi}(\boldsymbol{\xi}|\theta)\ \lambda_{2n}$-a.e., where ψ denotes the uniformly distributed random variable used for defining $\mathbf{x}'_{(\mathrm{a})}$ (see Definition 3.1), and, therefore, $\mathbf{x} = \mathbf{x}'_{(\mathrm{a})}\in\mathcal{I}$. □

Theorem 5.3: Suppose for (14) a random channel matrix $\mathbf{H}\in\mathbb{C}^{n\times m}$, a circular noise vector $\mathbf{z}\in\mathbb{C}^n$, and a circular-closed set

$\mathcal{I}$ of admissible input distributions. Then, there exists a circular random vector $\mathbf{x} \in \mathbb{C}^m$ that achieves the coherent capacity of (15).

Proof: Let us denote by $\mathbf{x}' \in \mathcal{I}$ a—not necessarily circular—random vector that achieves the coherent capacity of (15). Using the same line of arguments as in the proof of Theorem 5.1, its circular analog $\mathbf{x} \triangleq \mathbf{x}'_{(\mathrm{a})} = e^{j2\pi\psi}\mathbf{x}' \in \mathcal{I}$, where $\psi \in [0,1)$ is uniformly distributed and assumed to be independent of $\mathbf{x}'$, $\mathbf{z}$, and $\mathbf{H}$, can be shown to satisfy

$$I(\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}) \geq I(\mathbf{x}'^{(\mathrm{r})};\mathbf{y}'^{(\mathrm{r})}|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi})$$

where $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{z}$ and $\mathbf{y}' = \mathbf{H}\mathbf{x}'+\mathbf{z}$. It follows that

$$\begin{aligned} I(\mathbf{x};\mathbf{y},\mathbf{H}) &= \int I(\mathbf{x}^{(\mathrm{r})};\mathbf{y}^{(\mathrm{r})}|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \\ &\geq \int I(\mathbf{x}'^{(\mathrm{r})};\mathbf{y}'^{(\mathrm{r})}|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \\ &= I(\mathbf{x}';\mathbf{y}',\mathbf{H}) \\ &= C_{\mathrm{c}} \end{aligned}$$

i.e., $\mathbf{x}$ achieves the coherent capacity of (15).

B. Circular Channel Matrix

Here, we assume that the channel matrix $\mathbf{H} \in \mathbb{C}^{n\times m}$ is random, and—additionally—that an arbitrary stacking of the elements of $\mathbf{H}$ into an nm-dimensional vector yields a circular random vector. The noise vector $\mathbf{z}$ is not required to be circular. Note that this is the opposite situation compared with Section V-A, where $\mathbf{z}$ is circular but $\mathbf{H}$ is arbitrary. Again, it is assumed that the set $\mathcal{I}$ of admissible input distributions is circular-closed. For the input distributions that achieve the noncoherent capacity of (14) and the coherent capacity of (15), respectively, we have the following results.

Theorem 5.4: Suppose for (14) a random channel matrix $\mathbf{H} \in \mathbb{C}^{n\times m}$, such that the random vector, which is obtained from an arbitrary stacking of the elements of $\mathbf{H}$ into an nm-dimensional vector, is circular, and a circular-closed set $\mathcal{I}$ of admissible input distributions. Then, there exists a circular random vector $\mathbf{x} \in \mathbb{C}^m$ that achieves the noncoherent capacity of (14).

Proof: Let us denote by $\mathbf{x}' \in \mathcal{I}$ a—not necessarily circular—random vector that achieves the noncoherent capacity of (14), and let $\mathbf{x} \triangleq \mathbf{x}'_{(\mathrm{a})} = e^{j2\pi\psi}\mathbf{x}' \in \mathcal{I}$, where $\psi \in [0,1)$ is uniformly distributed and assumed to be independent of $\mathbf{x}'$, $\mathbf{z}$, and $\mathbf{H}$, be its circular analog. We have

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= h(\mathbf{H}\mathbf{x}+\mathbf{z}) - h(\mathbf{H}\mathbf{x}+\mathbf{z}|\mathbf{x}) \\ I(\mathbf{x}';\mathbf{y}') &= h(\mathbf{H}\mathbf{x}'+\mathbf{z}) - h(\mathbf{H}\mathbf{x}'+\mathbf{z}|\mathbf{x}') \end{aligned}$$

and intend to show $I(\mathbf{x};\mathbf{y}) = I(\mathbf{x}';\mathbf{y}')$. Due to the circularity of $\mathbf{H}$, Theorem 3.4 implies

$$h(\mathbf{H}\mathbf{x}+\mathbf{z}) = h\left(\mathbf{H}e^{j2\pi\psi}\mathbf{x}'+\mathbf{z}\right) = h(\mathbf{H}\mathbf{x}'+\mathbf{z})$$

so that it remains to show $h(\mathbf{H}\mathbf{x}+\mathbf{z}|\mathbf{x}) = h(\mathbf{H}\mathbf{x}'+\mathbf{z}|\mathbf{x}')$. Fubini's Theorem yields

$$\begin{aligned} &h(\mathbf{H}\mathbf{x}+\mathbf{z}|\mathbf{x}) \\ &= \int h\left(\mathbf{H}\mathbf{x}+\mathbf{z}\middle|\mathbf{x}^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= \int_0^1 \int h\left(\mathbf{H}\mathbf{x}+\mathbf{z}\middle|\mathbf{x}^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{(e^{j2\pi\varphi}\mathbf{x}')^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi}d\varphi \end{aligned}$$

since $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) = \int_0^1 f_{\mathbf{x}^{(\mathrm{r})}|\psi}(\boldsymbol{\xi}|\varphi)d\varphi = \int_0^1 f_{(e^{j2\pi\varphi}\mathbf{x}')^{(\mathrm{r})}}(\boldsymbol{\xi})d\varphi$, and, furthermore

$$\begin{aligned} &h(\mathbf{H}\mathbf{x}+\mathbf{z}|\mathbf{x}) \\ &= \int_0^1 \int h\left(\mathbf{H}e^{j2\pi\varphi}\mathbf{x}'+\mathbf{z}\middle|\mathbf{x}'^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}'^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi}d\varphi \\ &\overset{(*)}{=} \int_0^1 \int h\left(\mathbf{H}\mathbf{x}'+\mathbf{z}\middle|\mathbf{x}'^{(\mathrm{r})}=\boldsymbol{\xi}\right) f_{\mathbf{x}'^{(\mathrm{r})}}(\boldsymbol{\xi})d\boldsymbol{\xi}d\varphi \\ &= \int_0^1 h(\mathbf{H}\mathbf{x}'+\mathbf{z}|\mathbf{x}')\,d\varphi \\ &= h(\mathbf{H}\mathbf{x}'+\mathbf{z}|\mathbf{x}') \end{aligned}$$

where $(*)$ follows from the circularity of $\mathbf{H}$. Hence, the circular $\mathbf{x}$ achieves the noncoherent capacity of (14). □

Theorem 5.5: Suppose for (14) a random channel matrix $\mathbf{H} \in \mathbb{C}^{n\times m}$, such that the random vector, which is obtained from an arbitrary stacking of the elements of $\mathbf{H}$ into an nm-dimensional vector, is circular, and a circular-closed set $\mathcal{I}$ of admissible input distributions. Then, there exists a circular random vector $\mathbf{x} \in \mathbb{C}^m$ that achieves the coherent capacity of (15).

Proof: Let us denote by $\mathbf{x}' \in \mathcal{I}$ a—not necessarily circular—capacity-achieving random vector and let $\mathbf{x}_{(\theta)} \triangleq e^{j2\pi\theta}\mathbf{x}'$ (with $\theta \in [0,1)$ being deterministic). With $\mathbf{y}_{(\theta)} = \mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}$ we obtain

$$\begin{aligned} &I\left(\mathbf{x}_{(\theta)};\mathbf{y}_{(\theta)},\mathbf{H}\right) \\ &= \int I\left(\mathbf{x}_{(\theta)}^{(\mathrm{r})};\mathbf{y}_{(\theta)}^{(\mathrm{r})}\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \\ &= \int \Big(h\left(\mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right) \\ &\quad - h\left(\left(\mathbf{H}\mathbf{x}_{(\theta)}+\mathbf{z}|\mathbf{x}_{(\theta)}\right)\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right)\Big) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \\ &= \int \Big(h\left(\mathbf{H}e^{j2\pi\theta}\mathbf{x}'+\mathbf{z}\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right) \\ &\quad - h\left(\left(\mathbf{H}e^{j2\pi\theta}\mathbf{x}'+\mathbf{z}|\mathbf{x}'\right)\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right)\Big) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \\ &= \int \Big(h\left(\mathbf{H}\mathbf{x}'+\mathbf{z}\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right) \\ &\quad - h\left(\left(\mathbf{H}\mathbf{x}'+\mathbf{z}|\mathbf{x}'\right)\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right)\Big) f_{(e^{j2\pi\theta}\mathbf{H})^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \end{aligned}$$

where Fubini's Theorem has been used, and, furthermore, due to the circularity of $\mathbf{H}$

$$\begin{aligned} I\left(\mathbf{x}_{(\theta)};\mathbf{y}_{(\theta)},\mathbf{H}\right) &= \int I\left(\mathbf{x}'^{(\mathrm{r})};\mathbf{y}'^{(\mathrm{r})}\middle|\mathbf{H}^{(\mathrm{r})}=\boldsymbol{\chi}\right) f_{\mathbf{H}^{(\mathrm{r})}}(\boldsymbol{\chi})d\boldsymbol{\chi} \\ &= I(\mathbf{x}';\mathbf{y}',\mathbf{H}) \end{aligned}$$

where $\mathbf{y}' = \mathbf{H}\mathbf{x}'+\mathbf{z}$. Hence, $\mathbf{x}_{(\theta)}$ is capacity-achieving. Therefore, by applying Jensen's inequality to the concave mutual information function (with respect to the input distribution, cf., the proof of Theorem 5.2), the random vector $\mathbf{x} \in \mathbb{C}^m$ with distribution defined according to $f_{\mathbf{x}^{(\mathrm{r})}}(\boldsymbol{\xi}) \triangleq \int_0^1 f_{\mathbf{x}^{(\mathrm{r})}_{(\theta)}}(\boldsymbol{\xi}) d\theta$ λ_{2n}-a.e. satisfies

$$I(\mathbf{x};\mathbf{y},\mathbf{H}) \geq \int_0^1 I\left(\mathbf{x}_{(\theta)};\mathbf{y}_{(\theta)},\mathbf{H}\right) d\theta = \int_0^1 I(\mathbf{x}';\mathbf{y}',\mathbf{H})\, d\theta = I(\mathbf{x}';\mathbf{y}',\mathbf{H})$$

so that $\mathbf{x}$ achieves the coherent capacity of (15). But $f_{\mathbf{x}^{(\mathrm{r})}_{(\theta)}}(\boldsymbol{\xi}) = f_{\mathbf{x}'^{(\mathrm{r})}_{(\mathrm{a})}|\psi}(\boldsymbol{\xi}|\theta)$ λ_{2n}-a.e., where ψ denotes the uniformly distributed random variable used for defining $\mathbf{x}'_{(\mathrm{a})}$ (see Definition 3.1), and, therefore, $\mathbf{x} = \mathbf{x}'_{(\mathrm{a})} \in \mathcal{I}$. □

C. Deterministic Channel Matrix and Improper Gaussian Noise Vector

Here, we investigate the case that the channel matrix $\mathbf{H} \in \mathbb{C}^{n\times m}$ is deterministic and that the noise vector $\mathbf{z} \in \mathbb{C}^n$ is Gaussian distributed. We impose an average power constraint, i.e., we define the set of admissible input distributions as

$$\mathcal{I} \triangleq \{\mathbf{x} : \mathrm{E}\{\mathbf{x}^H\mathbf{x}\} \leq S\}. \tag{16}$$

For $\mathbf{z}$ proper, both capacity and capacity-achieving input vector are well known [15]. Therefore, in the following, we consider a more general situation without the assumption of $\mathbf{z}$ being proper. However, we introduce additional technical assumptions, which make the derivation less complicated and lead to simpler results. Note that most of these assumptions could be significantly relaxed or even omitted, but for the price of more involved theorems and proofs. We assume,

$$\mathbf{H} \in \mathbb{C}^{n\times n} \text{ deterministic, quadratic, and nonsingular} \tag{17a}$$

$$S \geq 2n\left\|\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right\|_2 \text{ (high signal}-\text{to}-\text{noise ratio)} \tag{17b}$$

$$\mathbf{z} \quad \text{zero-mean with nonsingular } \mathbf{C_z} \in \mathbb{C}^{n\times n} \tag{17c}$$

$$\left\|\mathbf{B_z}^{-1}\mathbf{P_z}\mathbf{B_z}^{-T}\right\|_2 < 1 \tag{17d}$$

where $\mathbf{C_z}$ and $\mathbf{P_z}$ denote covariance matrix and complementary covariance matrix of $\mathbf{z}$, respectively, and $\mathbf{B_z}$ is a generalized Cholesky factor of $\mathbf{C_z}$. We have the following capacity result.

Theorem 5.6: Suppose for (14) that assumptions (17) hold and that the set of admissible input distributions is defined according to (16). Then, the capacity of (14) is given by

$$C = 2\log|\det\mathbf{H}| + n\log\left(S + \mathrm{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right) - \log\det\mathbf{C_z} - \frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2) - n\log n$$

where λ_i are the singular values of $\mathbf{B_z}^{-1}\mathbf{P_z}\mathbf{B_z}^{-T}$. Furthermore, the zero-mean and Gaussian distributed random vector $\mathbf{x} \in \mathbb{C}^n$ with covariance matrix and complementary covariance matrix given by

$$\mathbf{C_x} = \frac{1}{n}\left(S + \mathrm{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right)\mathbf{I}_n - \mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H} \tag{18a}$$

$$\mathbf{P_x} = -\mathbf{H}^{-1}\mathbf{P_z}\mathbf{H}^{-T} \tag{18b}$$

respectively, is capacity-achieving.

Proof: Since $\mathrm{E}\{\mathbf{x}^H\mathbf{x}\} = \mathrm{tr}\,\mathbf{C_x} + \|\mathbf{m_x}\|_2^2$, where $\mathbf{m_x}$ denotes the mean vector of $\mathbf{x}$, Theorem 4.4 implies that the supremum of

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) = h(\mathbf{Hx}+\mathbf{z}) - h(\mathbf{z})$$

over $\mathcal{I}$ is achieved by a zero-mean and Gaussian distributed complex-valued random vector $\mathbf{x}$ with covariance matrix $\mathbf{C_x}$ that maximizes the function $g(\mathbf{C}) \triangleq \log\det\left(\mathbf{HCH}^H + \mathbf{C_z}\right)$ over the set of covariance matrices $\mathbf{C}$ with $\mathrm{tr}\,\mathbf{C} \leq S$, and with complementary covariance matrix $\mathbf{P_x}$ that satisfies $\mathbf{HP_xH}^T + \mathbf{P_z} = \mathbf{0}$, provided that such a random vector exists. Using the eigenvalue decomposition $\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H} = \mathbf{UDU}^H$ we obtain,

$$\begin{aligned} g(\mathbf{C}) &= 2\log|\det\mathbf{H}| + \log\det\left(\mathbf{C} + \mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right) \\ &= 2\log|\det\mathbf{H}| + \log\det\left(\mathbf{C} + \mathbf{UDU}^H\right) \\ &= 2\log|\det\mathbf{H}| + \log\det\left(\mathbf{U}^H\mathbf{CU} + \mathbf{D}\right) \end{aligned}$$

so that its maximum is achieved for $\mathbf{C} = \mathbf{C_x} \triangleq \mathbf{U}(L\mathbf{I}_n - \mathbf{D})\mathbf{U}^H = L\mathbf{I}_n - \mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}$, where L is chosen such that

$$S = \mathrm{tr}\,\mathbf{C_x} = Ln - \mathrm{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right) \tag{19}$$

is satisfied. Note that this is the well-known *water filling* solution [15], [34], with the additional simplification that $\mathbf{C_x}$ is nonsingular,[6] since, according to (17b)

$$\frac{S}{n} \geq 2\left\|\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right\|_2 = 2\|\mathbf{D}\|_2 = 2d_{\max} > d_{\max} \tag{20}$$

where $d_{\max}$ is the largest entry (eigenvalue) of $\mathbf{D}$. This yields (18a). Clearly, the choice (18b) satisfies $\mathbf{HP_xH}^T + \mathbf{P_z} = \mathbf{0}$. It remains to show that $\{\mathbf{C_x},\mathbf{P_x}\}$ is a valid pair of covariance matrix and complementary covariance matrix. To that end, consider

$$\begin{aligned} \|\mathbf{P_x}\|_2 &= \left\|\mathbf{H}^{-1}\mathbf{P_z}\mathbf{H}^{-T}\right\|_2 && \text{(21a)} \\ &= \left\|\mathbf{H}^{-1}\mathbf{B_z}\left(\mathbf{B_z}^{-1}\mathbf{P_z}\mathbf{B_z}^{-T}\right)\mathbf{B_z}^T\mathbf{H}^{-T}\right\|_2 \\ &\overset{(*)}{\leq} \left\|\mathbf{H}^{-1}\mathbf{B_z}\right\|_2^2 \\ &= \left\|\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right\|_2 \\ &= \|\mathbf{D}\|_2 \\ &= d_{\max} && \text{(21b)} \end{aligned}$$

where $(*)$ follows from Theorem 2.11, and note that $L > \frac{S}{n} \geq 2d_{\max}$, cf., (19) and (20). This implies

$$\|\mathbf{P_x}\|_2 < L - d_{\max} = \frac{1}{\left\|\mathbf{C_x}^{-1}\right\|_2}$$

and, furthermore

$$\left\|\mathbf{B_x}^{-1}\mathbf{P_x}\mathbf{B_x}^{-T}\right\|_2 \leq \left\|\mathbf{C_x}^{-1}\right\|_2\|\mathbf{P_x}\|_2 < 1$$

so that Theorem 2.11 shows that (18) defines a valid pair of covariance matrix and complementary covariance matrix. Finally, the capacity of (14) is obtained as

$$\begin{aligned} C &= g(\mathbf{C_x}) + n\log(\pi e) - h(\mathbf{z}) \\ &= 2\log|\det\mathbf{H}| + n\log\left(\frac{1}{n}\left(S + \mathrm{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right)\right) \end{aligned}$$

[6]The *water level* L is larger than the noise power for all (parallel) *eigenchannels*.

$$+ n\log(\pi e) - \log\det(\pi e \mathbf{C_z}) - \frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2)$$
$$= 2\log|\det \mathbf{H}| + n\log\left(S + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right)$$
$$- n\log n - \log\det \mathbf{C_z} - \frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2)$$

where Theorem 4.8 has been used. □

Remarks: Whereas in many real-world scenarios the noise vector happens to be circular, so that the results of Section V-A apply, there are also practically relevant scenarios, where the noise vector is known to be improper. More specifically, DMT modulation, which is widely used in *x*DSL applications [19], yields an equivalent system channel that exactly matches the situation considered here [9], [10]. We also note that capacity results for improper Gaussian distributed noise vectors could be alternatively derived by making use of an equivalent real-valued channel of dimension $2n \times 2m$ that is obtained by appropriate stacking of real and imaginary parts. The advantage of the approach presented here is that it yields expressions that are explicit in covariance matrix and complementary covariance matrix. In the following, we exploit this (desired) separation.

Observe that improper noise is beneficial since—due to $\frac{1}{2}\sum_{i=1}^{n}\log(1-\lambda_i^2) < 0$—it increases capacity. However, this presupposes a suitably designed transmission scheme. If it is erroneously believed that $\mathbf{P_z} = \mathbf{0}$, it will be erroneously believed as well (see Theorem 5.6) that the zero-mean and Gaussian distributed random vector $\mathbf{x}'$ with covariance matrix $\mathbf{C_{x'}} = \mathbf{C_x}$ as in (18a) but with $\mathbf{P_{x'}} = \mathbf{0}$ is capacity-achieving. It follows that

$$C' \stackrel{\triangle}{=} I(\mathbf{x}';\mathbf{y}') = C - \Delta C \le C$$

where $\mathbf{y}' = \mathbf{H}\mathbf{x}' + \mathbf{z}$ and $\Delta C \stackrel{\triangle}{=} C - C'$ denotes the resulting *capacity loss*. This capacity loss is quantified by the next theorem.

Theorem 5.7: Suppose for (14) that assumptions (17) hold and that the set of admissible input distributions is defined according to (16). Then, the capacity loss ΔC that occurs if it is erroneously believed that $\mathbf{P_z} = \mathbf{0}$ is given by

$$\Delta C = -\frac{1}{2}\sum_{i=1}^{n}\log(1-\mu_i^2)$$

where μ_i are the singular values of $\left(n/\left(S + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right)\right)\mathbf{H}^{-1}\mathbf{P_z}\mathbf{H}^{-T}$. In particular

$$0 \le \Delta C < n\log\frac{2}{\sqrt{3}}.$$

Proof: We intend to apply Theorem 4.8 to the random vector $\mathbf{y}' = \mathbf{H}\mathbf{x}' + \mathbf{z}$. In order to meet the assumption of Theorem 4.8, we have to show $\|\mathbf{B}_{\mathbf{y}'}^{-1}\mathbf{P}_{\mathbf{y}'}\mathbf{B}_{\mathbf{y}'}^{-T}\|_2 < 1$, where $\mathbf{B}_{\mathbf{y}'}$ denotes a generalized Cholesky factor of $\mathbf{C}_{\mathbf{y}'}$. Note that the nonsingularity of $\mathbf{C}_{\mathbf{y}'}$ follows from the nonsingularity of $\mathbf{C_z}$. Clearly, $\mathbf{C}_{\mathbf{y}'} = \mathbf{H}\mathbf{C_x}\mathbf{H}^H + \mathbf{C_z} = \frac{1}{n}\left(S + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right)\mathbf{H}\mathbf{H}^H$ and $\mathbf{P}_{\mathbf{y}'} = \mathbf{P_z}$, so that $\mathbf{B}_{\mathbf{y}'}^{-1}\mathbf{P}_{\mathbf{y}'}\mathbf{B}_{\mathbf{y}'}^{-T} = \left(n/\left(S + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)\right)\right)\mathbf{H}^{-1}\mathbf{P_z}\mathbf{H}^{-T}$ and, furthermore

$$\begin{aligned}\|\mathbf{B}_{\mathbf{y}'}^{-1}\mathbf{P}_{\mathbf{y}'}\mathbf{B}_{\mathbf{y}'}^{-T}\|_2 &= \frac{n}{S + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)} \times \\ &\qquad \times \|\mathbf{H}^{-1}\mathbf{P_z}\mathbf{H}^{-T}\|_2 \\ &\stackrel{(21a),(21b)}{\le} \frac{n}{S + \operatorname{tr}\left(\mathbf{H}^{-1}\mathbf{C_z}\mathbf{H}^{-H}\right)}\, d_{\max} \\ &< \frac{n}{S}\, d_{\max} \\ &\stackrel{(20)}{\le} \frac{1}{2}.\end{aligned} \tag{22}$$

The capacity loss is then given by

$$\begin{aligned}\Delta C &= C - C' \\ &= h(\mathbf{y}) - h(\mathbf{z}) - h(\mathbf{y}') + h(\mathbf{z}) \\ &= h(\mathbf{y}) - h(\mathbf{y}') \\ &\stackrel{(*)}{=} \log\det(\pi e\mathbf{C_y}) - \log\det(\pi e\mathbf{C_y}) - \frac{1}{2}\sum_{i=1}^{n}\log(1-\mu_i^2) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\log(1-\mu_i^2)\end{aligned}$$

where $(*)$ follows from Theorem 4.8, since $\mathbf{P_y} = \mathbf{0}$ and $\mathbf{C}_{\mathbf{y}'} = \mathbf{C_y}$. For the bound note that

$$\begin{aligned}-\frac{1}{2}\sum_{i=1}^{n}\log(1-\mu_i^2) &\le -\frac{n}{2}\log(1 - \|\mathbf{B}_{\mathbf{y}'}^{-1}\mathbf{P}_{\mathbf{y}'}\mathbf{B}_{\mathbf{y}'}^{-T}\|_2^2) \\ &\stackrel{(22)}{<} -\frac{n}{2}\log\left(1-\frac{1}{4}\right) \\ &= n\log\frac{2}{\sqrt{3}}.\end{aligned}$$

□

Example: Let us consider the special case of the complex scalar channel $y = x + z$ with noise covariance $C_z \in \mathbb{R}$ and complementary noise covariance $P_z \in \mathbb{R}$, where $C_z \ge P_z > 0$. According to (3)

$$\mathbf{C}_{z^{(\mathrm{r})}} = \frac{1}{2}\begin{bmatrix} C_z + P_z & 0 \\ 0 & C_z - P_z \end{bmatrix}$$

which is illustrated in Fig. 1(a). It is seen that the noise power is different for real and imaginary part. If it is erroneously believed that $P_z = 0$, the same power will be assigned to real and imaginary part of the input vector, as it is shown in Fig. 1(b). However, the optimum power distribution that maximizes the mutual information is different; it is depicted in Fig. 1(c). Note that this capacity-achieving power distribution is obtained by water filling on a real and imaginary part level. The difference between the mutual informations of solutions (b) and (c) is expressed by the capacity loss ΔC.

VI. Conclusion

We studied the influence of circularity/noncircularity and properness/improperness on important information theoretic quantities such as entropy, divergence, and capacity. As a motivating starting point served a theorem by Neeser and

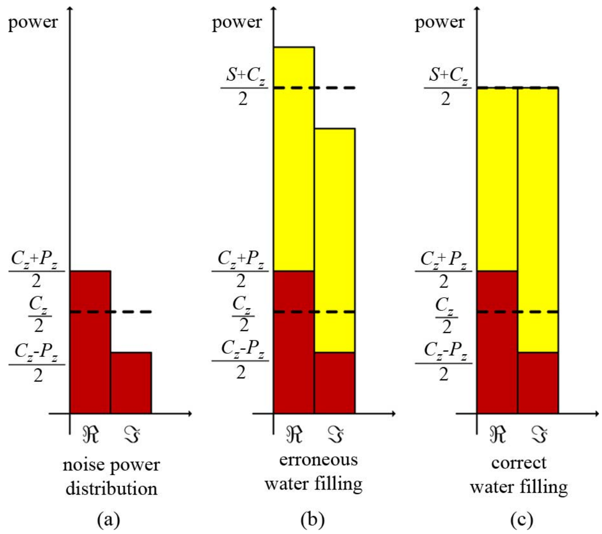

Fig. 1. Water filling strategies illustrating the capacity loss.

Massey [14], which states that the entropy of a zero-mean complex-valued random vector is upper-bounded by the entropy of a circular/proper Gaussian distributed random vector with same covariance matrix. We strengthened this theorem in two different directions: 1) we dropped the Gaussian assumption and 2) we dropped the properness assumption. In both cases, the resulting upper bound turned out to be tighter than the one previously known. A key ingredient for the proof in case 1 was the introduction of the *circular analog* of a given complex-valued random vector. Whereas its definition was based on intuitive arguments to obtain a circular random vector, which is "close" to the (potentially) noncircular given one, we rigorously proved that it equals the unique circular random vector with minimum Kullback–Leibler divergence. On the other hand, for case 2, we exploited results about the second-order structure of complex-valued random vectors that were obtained without making use of the augmented covariance matrix (in contrast to related work). Additionally, we presented a criterion for a matrix to be a valid complementary covariance matrix. Furthermore, we addressed the capacity of MIMO channels. Regardless of the specific distribution of the channel parameters (noise vector and channel matrix, if modeled as random), we showed that the capacity-achieving input vector is circular for a broad range of MIMO channels (including coherent and noncoherent scenarios). This extends known results that make use of a Gaussian assumption. Finally, we investigated the situation of an improper and Gaussian distributed noise vector. We computed both capacity and capacity-achieving input vector and showed that improperness increases capacity, provided that the complementary covariance matrix is exploited. Otherwise, a capacity loss occurs, for which we derived an explicit expression.

Acknowledgment

The author would like to thank J. Huber, J. Sayir, and J. Weinrichter for helpful hints and comments. He is also grateful to the anonymous reviewers for their constructive comments that have resulted in a major improvement of this paper.

References

[1] G. Tauböck, "Rotationally variant complex channels," in *Proc. 23rd Symp. Inf. Theory Benelux*, Louvain-la-Neuve, Belgium, May 2002, pp. 261–268.

[2] G. Tauböck, "On the maximum entropy theorem for complex random vectors," in *Proc. IEEE Int. Symp. Inf. Theory*, IL, Chicago, Jun./Jul. 2004, p. 41.

[3] G. Tauböck, "A maximum entropy theorem for complex-valued random vectors, with implications on capacity," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 375–379.

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

[5] Y. C. Yoon and H. Leib, "Maximizing SNR in improper complex noise and applications to CDMA," *IEEE Commun. Lett.*, vol. 1, no. 1, pp. 5–8, Jan. 1997.

[6] 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.

[7] A. Lampe, R. Schober, W. Gerstacker, and J. Huber, "A novel iterative multiuser detector for complex modulation schemes," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 2, pp. 339–350, Feb. 2002.

[8] H. Gerstacker, R. Schober, and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1512–1523, Sep. 2003.

[9] G. Tauböck, "Noise analysis of DMT," in *Proc. IEEE Global Commun. Conf.*, San Francisco, CA, Dec. 2003, vol. 4, pp. 2136–2140.

[10] G. Tauböck, "Complex noise analysis of DMT," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5739–5754, Dec. 2007.

[11] L. Anttila, M. Valkama, and M. Renfors, "Circularity-based I/Q imbalance compensation in wideband direct-conversion receivers," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2099–2113, Jul. 2008.

[12] P. Rykaczewski, M. Valkama, and M. Renfors, "On the connection of I/Q imbalance and channel equalization in direct-conversion transceivers," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1630–1636, May 2008.

[13] Y. Zou, M. Valkama, and M. Renfors, "Digital compensation of I/Q imbalance effects in space-time coded transmit diversity systems," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2496–2508, Jun. 2008.

[14] 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.

[15] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[16] P. J. Schreier, L. L. Scharf, and C. T. Mullis, "Detection and estimation of improper complex random signals," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 306–312, Jan. 2005.

[17] P. J. Schreier, L. L. Scharf, and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 433–436, Jul. 2006.

[18] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1017–1029, Mar. 2006.

[19] M. Gagnaire, "An overview of broad-band access technologies," *Proc. IEEE*, vol. 85, no. 12, pp. 1958–1972, Dec. 1997.

[20] P. R. Halmos, *Measure Theory*. New York: Springer-Verlag, 1974.

[21] G. Tauböck, "Wireline multiple-input/multiple-output systems," Ph.D. dissertation, Inst. Commun. Radio-Frequency Eng., Vienna Univ. Technology, Vienna, Austria, 2005.

[22] W. Rudin, *Real and Complex Analysis 3rd International Edition*. New York: McGraw-Hill, 1987.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[25] K. Karhunen, "Über lineare Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae, ser. A (I)*, vol. 37, no. 18, pp. 3–79, 1947.

[26] M. Loève, "Fonctions Aléatoires du Second Ordre," in *Processus stochastiques et mouvements Browniens*, P. Lévy, Ed. Paris, France: Gauthier-Villars, 1948.

[27] H. Hotelling, "Analysis of a complex pair of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, 1933.

[28] Y. Eldar and A. Oppenheim, "MMSE whitening and subspace whitening," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1846–1851, Jul. 2003.

[29] P. J. Schreier and L. L. Scharf, "Canonical coordinates for reduced-rank estimation of improper complex random vectors," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Orlando, FL, May 2002, pp. 1153–1156.

[30] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3–4, pp. 321–377, 1936.

[31] L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Process.*, vol. 36, no. 3, pp. 647–654, Mar. 1998.

[32] L. L. Scharf and C. T. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 824–831, Mar. 2000.

[33] A. Papoulis*, Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[34] T. M. Cover and J. A. Thomas*, Elements of Information Theory*. New York: Wiley, 1991.

[35] A. Dembo and O. Zeitouni*, Large Deviations Techniques and Applications (corrected printing of the 1998 edition)*. Berlin, Germany: Springer-Verlag, 2010.

[36] M. Abramowitz and I. Stegun*, Handbook of Mathematical Functions*. New York: Dover, 1965.

[37] R. M. Gray*, Entropy and Information Theory*, 2nd ed. New York: Springer-Verlag, 2010.

[38] R. B. Ash*, Probability and Measure Theory*, 2nd ed. San Diego, CA: Academic, 2000.

Georg Tauböck (S'01–M'07) received the Dipl.-Ing. degree and the Dr.techn. degree (with highest honors) in electrical engineering and the Dipl.-Ing. degree in mathematics (with highest honors) from Vienna University of Technology, Vienna, Austria in 1999, 2005, and 2008, respectively. He also received the diploma in violoncello from the Conservatory of Vienna, Vienna, Austria, in 2000.

From 1999 to 2005, he was with the FTW Telecommunications Research Center Vienna, Vienna, Austria, and since 2005, he has been with the Institute of Telecommunications, Vienna University of Technology, Vienna, Austria. From February 2010 to August 2010, he was a visiting researcher with the Communication Technology Laboratory (Communication Theory Group) at ETH Zurich, Zurich, Switzerland.

His research interests include wireline and wireless communications, compressed sensing, signal processing, and information theory.