

A Maximum Entropy Theorem for Complex-Valued Random Vectors, with Implications on Capacity

Georg Tauböck

Institute of Telecommunications, Vienna University of Technology, A-1040 Vienna, Austria

Email: gtauboc@nt.tuwien.ac.at

Abstract—Recent research has demonstrated significant achievable performance gains by exploiting circularity/non-circularity 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 and capacity. More specifically, we prove a novel maximum entropy theorem that is based on the so-called *circular analog* of a given (in general, non-Gaussian) complex-valued random vector. Its introduction is supported by a characterization theorem that employs a minimum Kullback-Leibler divergence criterion. As an application of this maximum entropy theorem, we show that the capacity-achieving input random vector is circular for a broad range of multiple-input multiple-output (MIMO) channels including coherent and noncoherent scenarios. This result does not depend on a Gaussian assumption and thus provides a justification for many practical signalling/coding strategies, regardless of the specific distribution of the channel parameters.

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 an comprehensive overview see [1]—it has been shown that exploiting *circularity* or its second-order counterpart *properness* of complex-valued signals or lack of it (*non-circularity/improperness*) is able to significantly enhance the performance of signal processing techniques. However, there are only a few information theoretic results in this field [1–4] 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 result that addresses complex-valued random vectors analyzes the differential entropy of complex-valued random vectors [2]. 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 [1]) *Gaussian* distributed complex-valued random vector with the same covariance matrix. Note that for a *non-Gaussian* random vector, this upper bound is not tight.

In order to derive a maximum entropy theorem with a tight upper bound, we have to deal with the problem, how we can

This work was supported by WWTF grant SPORTS (MA 07-004) and by FWF grant “Statistical Inference” (S10603-N13) within the National Research Network SISE.

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—in the following termed *circular analog* of a complex-valued random vector—is intuitive, and is furthermore supported by a characterization theorem that is based on a minimum Kullback-Leibler divergence criterion. As we will show, this will lead to an improved maximum entropy theorem, the central result of this paper.

Finally, we apply the obtained theorem to investigate structural properties of the capacity-achieving input distribution for a broad range of complex-valued MIMO channels (without a Gaussian assumption). In particular, we show that circularity of either the channel matrix or the noise vector is sufficient for circularity of the capacity-achieving input vector. Although this result seems to be quite intuitive, it was previously only known for the special case of a circular *Gaussian* noise vector [5]. Furthermore, contrary to [5], the proof techniques employed here require non-trivial tools from measure theory. **Notation.** The $n \times n$ identity matrix is denoted by \mathbf{I}_n . We use the superscript $[\cdot]^T$ for transposition and the superscript $[\cdot]^H \triangleq ([\cdot]^T)^*$ 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 part, respectively, and $\mathbb{E}\{\cdot\}$ refers to usual expectation. Throughout the paper, $\log(\cdot)$ denotes the logarithm taken with respect to an arbitrary but fixed base.

Outline. The remainder of this paper is organized as follows. In Section II, we introduce the framework and present initial results about the distribution 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 the novel maximum entropy theorem is derived. Finally, in Section V, we present various results about the capacity-achieving input vector of complex-valued MIMO channels under certain circularity assumptions. Due to lack of space, we have omitted the proofs, which are included in the full paper [6, available online].

II. FRAMEWORK AND PRELIMINARY RESULTS

We consider complex-valued random vectors $\mathbf{x} \in \mathbb{C}^n$. We assume that $\mathbf{x}^{(r)} \in \mathbb{R}^{2n}$, where $\mathbf{x}^{(r)} \triangleq [\Re\{\mathbf{x}^T\} \Im\{\mathbf{x}^T\}]^T$ 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) $f_{\mathbf{x}^{(r)}}(\boldsymbol{\xi})$. More precisely, it is assumed that the measure¹ defining the distribution of $\mathbf{x}^{(r)}$ is absolutely continuous with respect to λ_{2n} , where λ_{2n} denotes the $2n$ -dimensional Lebesgue measure [7]. 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 $\mathbf{x}^{(r)}$ defined by the pdf $f_{\mathbf{x}^{(r)}}(\boldsymbol{\xi})$. Hence, a complex-valued random vector \mathbf{x} will be called Gaussian distributed if $\mathbf{x}^{(r)}$ 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 *non-circular*. 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 . ■

Note that circularity implies properness (under the assumption of existing first- and second-order moments) and a vanishing mean vector. Conversely, a zero-mean, proper and Gaussian random vector is circular, but this does not hold for arbitrary distributions [1].

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

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

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

Lemma 2.2: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector, which is distributed according to the pdf $f_{\mathbf{x}^{(r)}}(\boldsymbol{\xi})$. Then, the pdf of its polar representation $\mathbf{x}^{(p)}$ is given by

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

almost everywhere with respect to λ_{2n} (λ_{2n} -a.e.) [7]. ■

Observing that the pdf of $\mathbf{y}_{(\theta)}^{(p)}$ of the random vector $\mathbf{y}_{(\theta)} \triangleq e^{j2\pi\theta}\mathbf{x}$ (with $\theta \in [0, 1[$ being deterministic) satisfies $f_{\mathbf{y}_{(\theta)}^{(p)}}(r_1, \dots, r_n, \phi_1, \dots, \phi_n) = f_{\mathbf{x}^{(p)}}(r_1, \dots, r_n, [\phi_1 - \theta]_{[0, 1[}, \dots, [\phi_n - \theta]_{[0, 1[})$ 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.

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

Corollary 2.3: A complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is circular if and only if the pdf of its *polar representation* $\mathbf{x}^{(p)}$ satisfies

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

■

Let us denote by $\mathbb{T}^{(s \rightarrow p)}$ the mapping

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

which is one-to-one with inverse $\mathbb{T}^{(p \rightarrow s)} \triangleq (\mathbb{T}^{(s \rightarrow p)})^{-1}$ given by

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

This follows immediately from the identity

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

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

Lemma 2.4: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the pdfs of its polar representation $\mathbf{x}^{(p)}$ and its sheared-polar representation $\mathbf{x}^{(s)}$ are related according to

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

■

Combining Corollary 2.3 and Lemma 2.4, while applying (1), yields the following important corollary.

Corollary 2.5: A complex-valued random vector $\mathbf{x} \in \mathbb{C}^n$ is circular if and only if the pdf of its *sheared-polar representation* $\mathbf{x}^{(s)}$ does not depend on ϕ_n , i.e., if and only if

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

Remark. While Lemma 2.2 is a direct consequence of the change-of-variables theorem [8], the proof of Lemma 2.4 and, consequently, Corollary 2.5 is rather sophisticated due to the non-continuity of $\mathbb{T}^{(s \rightarrow p)}$, cf. the full paper [6] for details.

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 non-circular. 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) non-circular one is motivated by the well-known method used for stationarizing a cyclostationary random process [9].

Definition 3.1: Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the random vector $\mathbf{x}_{(a)} \triangleq e^{j2\pi\psi} \mathbf{x}$, 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 $\mathbf{x}_{(a)}$ 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 $\mathbf{x}_{(a)}$ are related according to

$$\begin{aligned} f_{\mathbf{x}_{(a)}^{(p)}}(r_1, \dots, r_n, \phi_1, \dots, \phi_n) &= \int_0^1 f_{\mathbf{x}^{(p)}}(r_1, \dots, r_n, [\phi_1 - \varphi]_{[0,1]}, \dots, [\phi_n - \varphi]_{[0,1]}) d\varphi \quad \lambda_{2n}\text{-a.e.}, \\ f_{\mathbf{x}_{(a)}^{(s)}}(r_1, \dots, r_n, \phi_1, \dots, \phi_n) &= \int_0^1 f_{\mathbf{x}^{(s)}}(r_1, \dots, r_n, \phi_1, \dots, \phi_n) d\phi_n \quad \lambda_{2n}\text{-a.e.}, \end{aligned} \quad (2)$$

respectively. ■

Observe that $f_{\mathbf{x}_{(a)}^{(s)}}$ does not depend on ϕ_n λ_{2n} -a.e., so that Corollary 2.5 implies circularity of $\mathbf{x}_{(a)}$.

Next, we present a (non-trivial) 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*) [10, 11], 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

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

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 $f_{\mathbf{x}^{(r)}}$ is contained in the support set of $f_{\mathbf{y}^{(r)}}$ λ_{2n} -a.e. Note that $D(\mathbf{x}||\mathbf{y}) = 0$ if and only if $f_{\mathbf{x}^{(r)}} = f_{\mathbf{y}^{(r)}}$ λ_{2n} -a.e. [10]. 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

$$D(\mathbf{x}||\mathbf{y}) = D(\mathbf{x}^{(p)}||\mathbf{y}^{(p)}) = \int_{(\mathbb{R}_0^+)^n \times ([0,1]^n)} f_{\mathbf{x}^{(p)}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(p)}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(p)}}(\boldsymbol{\xi})} d\boldsymbol{\xi}$$

$$= D(\mathbf{x}^{(s)}||\mathbf{y}^{(s)}) = \int_{(\mathbb{R}_0^+)^n \times ([0,1]^n)} f_{\mathbf{x}^{(s)}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x}^{(s)}}(\boldsymbol{\xi})}{f_{\mathbf{y}^{(s)}}(\boldsymbol{\xi})} d\boldsymbol{\xi}. \quad \blacksquare$$

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., $\mathbf{x}^{(s)} \in \mathbb{R}^{2n}$, and form the “reduced” vector $\tilde{\mathbf{x}}^{(s)} \in \mathbb{R}^{2n-1}$ by only taking the first $2n - 1$ elements of $\mathbf{x}^{(s)}$. Clearly, its pdf is given by marginalization, i.e., $f_{\tilde{\mathbf{x}}^{(s)}}(\tilde{\boldsymbol{\xi}}) = \int_0^1 f_{\mathbf{x}^{(s)}}(r_1, \dots, r_n, \phi_1, \dots, \phi_n) d\phi_n$, where $\tilde{\boldsymbol{\xi}} \triangleq (r_1, \dots, r_n, \phi_1, \dots, \phi_{n-1})$. Furthermore, let $\tilde{\mathcal{S}}_{\mathbf{x}} \subset (\mathbb{R}_0^+)^n \times ([0, 1]^{n-1})$ denote the support set of $f_{\tilde{\mathbf{x}}^{(s)}}$. Note that $f_{\tilde{\mathbf{x}}^{(s)}}(\tilde{\boldsymbol{\xi}}) = 0$ is equivalent to $f_{\mathbf{x}^{(s)}}(\boldsymbol{\xi}, \phi_n) = 0$ λ_1 -a.e. (for fixed $\tilde{\boldsymbol{\xi}}$). We have,

$$f_{\mathbf{x}^{(s)}}(\tilde{\boldsymbol{\xi}}, \phi_n) = f_{\tilde{\mathbf{x}}^{(s)}}(\tilde{\boldsymbol{\xi}}) f_{\vartheta|\tilde{\mathbf{x}}^{(s)}}(\phi_n|\tilde{\boldsymbol{\xi}}), \quad \tilde{\boldsymbol{\xi}} \in \tilde{\mathcal{S}}_{\mathbf{x}}, \quad \phi_n \in [0, 1[, \quad (4)$$

where $\vartheta \triangleq (\mathbf{x}^{(s)})_{2n}$ denotes the last element of $\mathbf{x}^{(s)}$.

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., $\mathbf{y} = \mathbf{x}_{(a)}$, 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

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

Furthermore,

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

where $h(\vartheta|\tilde{\mathbf{x}}^{(s)})$ denotes the *conditional differential entropy* of ϑ given $\tilde{\mathbf{x}}^{(s)}$, cf. [10] and Definition 4.2, with ϑ and $\tilde{\mathbf{x}}^{(s)}$ according to (4). ■

Suppose \mathbf{x} is a complex-valued random vector, whose second-order moments exist. Clearly, both mean vector and complementary covariance matrix [1] of $\mathbf{x}_{(a)}$ 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 $\mathbf{x}_{(a)}$ equals the covariance matrix of \mathbf{x} , i.e., $\mathbf{C}_{\mathbf{x}_{(a)}} = \mathbf{C}_{\mathbf{x}}$. ■

IV. DIFFERENTIAL ENTROPY OF COMPLEX-VALUED RANDOM VECTORS

As outlined in the introduction, 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 $\mathbf{x}^{(r)}$, i.e.,

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

provided that the integrand is integrable [7]. ■

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 $\mathbf{x}^{(r)}$ given the real representation $\mathbf{y}^{(r)}$, i.e.,

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

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

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 $\mathbf{x}^{(r)}$ and $\mathbf{y}^{(r)}$, i.e.,

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

where $f_{\mathbf{x}^{(r)},\mathbf{y}^{(r)}}(\boldsymbol{\xi}, \boldsymbol{\eta})$ denotes the joint pdf of $\mathbf{x}^{(r)}$ and $\mathbf{y}^{(r)}$, and $f_{\mathbf{x}^{(r)}}(\boldsymbol{\xi})$ and $f_{\mathbf{y}^{(r)}}(\boldsymbol{\eta})$ are the marginal pdfs of $\mathbf{x}^{(r)}$ and $\mathbf{y}^{(r)}$, respectively. ■

It is well known that these quantities satisfy the following relations,

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}), \quad (5a)$$

$$I(\mathbf{x}; \mathbf{y}) \geq 0, \quad (5b)$$

with equality in (5b) 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 (see e.g., [2, 5]).

Theorem 4.4 (Neeser & Massey): Suppose $\mathbf{x} \in \mathbb{C}^n$ is a zero-mean complex-valued random vector with non-singular covariance matrix $\mathbf{C}_{\mathbf{x}}$. Then, the differential entropy of \mathbf{x} satisfies

$$h(\mathbf{x}) \leq \log \det(\pi e \mathbf{C}_{\mathbf{x}}), \quad (6)$$

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

Let us assume, for the moment, that \mathbf{x} is known to be non-Gaussian. Clearly, the inequality (6) is strict in this case and $\log \det(\pi e \mathbf{C}_{\mathbf{x}})$ is not a tight upper bound for the differential entropy $h(\mathbf{x})$. In the following, we will derive an improved maximum entropy theorem that takes this observation into

account. While its application is not limited to the non-Gaussian case, the obtained upper bound is in general tighter than the upper bound given by 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) Suppose $\mathbf{x} \in \mathbb{C}^n$ is a complex-valued random vector. Then, the differential entropies of \mathbf{x} and its circular analog $\mathbf{x}_{(a)}$ satisfy

$$h(\mathbf{x}) \leq h(\mathbf{x}_{(a)}),$$

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

Remarks. Since $\mathbf{x}_{(a)}$ is non-Gaussian in general,² the upper bound in Theorem 4.5 is typically tighter than the upper bound in Theorem 4.4, as can be seen by applying Theorem 3.5. Furthermore, Theorem 4.5 does not need the requirement of finite second-order moments.

V. CAPACITY-ACHIEVING INPUT DISTRIBUTION OF CIRCULAR MIMO CHANNELS

In this section we study the influence of circularity on channel capacity. More specifically, we investigate structural properties of the capacity-achieving input distribution. We consider vector-valued (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

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (7)$$

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. Both³ \mathbf{x} and \mathbf{z} are modeled as iid (only with respect to channel uses; within the random vectors the iid assumption is not made) vector-valued random processes, whereas \mathbf{H} is either assumed to be deterministic or is modeled as an iid (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 we do *not* assume that \mathbf{z} or \mathbf{H} (if applicable) are Gaussian distributed. The channel is characterized by the conditional distribution of \mathbf{y} given \mathbf{x} via the conditional pdf $f_{\mathbf{y}^{(r)}|\mathbf{x}^{(r)}}(\boldsymbol{\eta}|\boldsymbol{\xi})$ of their real representations $\mathbf{y}^{(r)}$ given $\mathbf{x}^{(r)}$, 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 $f_{\mathbf{x}^{(r)}}$ is in \mathcal{I} . Then, the *capacity/noncoherent capacity* of (7) is given by the supremum of the mutual information over the set of admissible input distributions [12], i.e., by

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

²Note that it is possible to define an improper (non-Gaussian) 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.

³Without loss of generality, since an iid (with respect to channel uses) \mathbf{x} is capacity achieving if \mathbf{z} and \mathbf{H} (if applicable) are iid.

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 (7) is the pair

$$(\mathbf{y}, \mathbf{H}) = (\mathbf{H}\mathbf{x} + \mathbf{z}, \mathbf{H}), \quad (8)$$

so that the channel law of (8) is governed by the conditional pdf $f_{\mathbf{y}^{(r)}; \mathbf{H}^{(r)} | \mathbf{x}^{(r)}}(\boldsymbol{\eta}, \boldsymbol{\chi} | \boldsymbol{\xi})$, where $\mathbf{H}^{(r)}$ is defined by an appropriate stacking of real and imaginary part of \mathbf{H} . Therefore, the *coherent capacity* of (8) is given by

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

where $f_{\mathbf{H}^{(r)}}(\boldsymbol{\chi})$ denotes the pdf of $\mathbf{H}^{(r)}$ and Fubini's Theorem has been used. A random vector $\mathbf{x} \in \mathcal{I}$ is said to be *capacity-achieving* for (7) or (8), if $I(\mathbf{x}; \mathbf{y}) = C$ or $I(\mathbf{x}; \mathbf{y}, \mathbf{H}) = C_c$, respectively.

In what follows, we assume that \mathcal{I} contains a capacity-achieving input vector and that \mathcal{I} is closed under the operation of forming the circular analog, i.e., that $\mathbf{x} \in \mathcal{I}$ implies $\mathbf{x}_{(a)} \in \mathcal{I}$ —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 peak and average power constraints, cf. Theorem 3.5.

A. Circular Noise Vector

Here, we assume that the noise vector $\mathbf{z} \in \mathbb{C}^n$ is circular. For a Gaussian distributed \mathbf{z} it has been shown in [5] that capacity (for deterministic \mathbf{H}) and coherent capacity (for random \mathbf{H}) are achieved by circular (Gaussian distributed) random vectors, respectively. The following theorems extend these results to the non-Gaussian case.

Theorem 5.1: Suppose for (7) 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 (7). ■

Theorem 5.2: Suppose for (7) 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 (7). ■

Theorem 5.3: Suppose for (7) 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 (8). ■

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. We have the following results about the capacity-achieving input vector.

Theorem 5.4: Suppose for (7) 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 (7). ■

Theorem 5.5: Suppose for (7) 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 (8). ■

VI. CONCLUSION

We studied the influence of circularity and non-circularity on important information theoretic quantities such as entropy and capacity. We presented a maximum entropy theorem, whose upper-bound improves upon existing results. A key ingredient for the proof was the introduction of the *circular analog* of a given complex-valued random vector, which was shown to equal the unique circular random vector with minimum Kullback-Leibler divergence. 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.

REFERENCES

- [1] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge (UK): Cambridge Univ. Press, 2010.
- [2] 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, July 1993.
- [3] G. Tauböck, "Rotationally variant complex channels," in *Proc. 23rd Symp. Inf. Theory Benelux*, Louvain-la-Neuve, Belgium, May 2002, pp. 261–268.
- [4] —, "On the maximum entropy theorem for complex random vectors," in *Proc. IEEE ISIT-2004*, Chicago, IL, p. 41, Jun./Jul. 2004.
- [5] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [6] G. Tauböck, "Complex-valued random vectors and channels: Entropy, divergence, and capacity," *IEEE Trans. Inf. Theory*, to appear (available online: <http://arxiv.org/abs/1105.0769>).
- [7] P. R. Halmos, *Measure Theory*. New York: Springer-Verlag, 1974.
- [8] W. Rudin, *Real and Complex Analysis (3rd International Edition)*. New York: McGraw Hill, 1987.
- [9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [11] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications (corrected printing of the 1998 edition)*. Berlin Heidelberg: Springer-Verlag, 2010.
- [12] R. M. Gray, *Entropy and Information Theory (2nd Edition)*. New York: Springer-Verlag (available online: <http://ee.stanford.edu/~gray/it.pdf>), 2010.