Appendix A

A.1 Appendix-I: Complex Gaussian Random Vectors

For any $z \in \mathbb{C}^n$ and $A \in \mathbb{C}^{n \times m}$, let us define

$$\hat{z} = \begin{pmatrix} \mathbb{R}e\,(z) \\ \mathbb{I}m(z) \end{pmatrix}$$

and

$$\hat{A} = \begin{pmatrix} \mathbb{R}e\,(A) & -\mathbb{I}m(A) \\ \mathbb{I}m(A) & \mathbb{R}e\,(A) \end{pmatrix}.$$

A complex random vector $\xi \in \mathbb{C}^n$ is said to be Gaussian if the real random vector $\hat{\xi} \in \mathbb{R}^{2n}$ consisting of the real and imaginary parts of ξ,

$$\hat{\xi} = \begin{pmatrix} \mathbb{R}e\,(\xi) \\ \mathbb{I}m\,(\xi) \end{pmatrix},$$

is Gaussian [1]. In fact, any complex random vector, without restricting to a complex Gaussian random vector, has such a real and expanded random vector. The relationship between these two complex and real vectors is a one to one mapping. Let us recall

$$E\left[\hat{\xi}\right] \triangleq \int_{\mathbb{R}^{2n}} x dP_{\hat{\xi}}\,(x) \in \mathbb{R}^{2n}$$

and

$$E\left[\left(\hat{\xi} - E\left[\hat{\xi}\right]\right)\left(\hat{\xi} - E\left[\hat{\xi}\right]\right)^T\right] \triangleq \int_{\mathbb{R}^{2n}} \left(x - E\left[\hat{\xi}\right]\right)\left(x - E\left[\hat{\xi}\right]\right)^T dP_{\hat{\xi}}\,(x)$$

P. He et al., *Radio Resource Management Using Geometric Water-Filling*, SpringerBriefs in Computer Science, DOI 10.1007/978-3-319-04636-5, © The Author(s) 2014

$\in \mathbb{R}^{2n\times 2n}$, the former is called the mean of $\hat{\xi}$ and the latter is called the covariance of $\hat{\xi}$. Note that $\mathbb{R}^{2n\times 2n}$ stands for the set of real square matrices with the size of $(2n) \times (2n)$, i.e., $2n$ rows and $2n$ columns. According to standard Lebesgue integration [2] on $\mathbb{R}^{2n}$, the mean and covariance of $\hat{\xi}$ can be found respectively. Thus, to specify the distribution of a complex Gaussian random vector ξ, it is necessary to specify the mean and covariance of $\hat{\xi}$, namely,

$$E\left[\hat{\xi}\right] \text{ and } E\left[\left(\hat{\xi}-E\left[\hat{\xi}\right]\right)\left(\hat{\xi}-E\left[\hat{\xi}\right]\right)^T\right].$$

The definitions of the mean and covariance are also suitable for the case of any complex random vector.

According to standard Lebesgue integration on $\mathbb{C}^n$, mean μ and covariance Q of ξ can be defined as follows.

$$\mu = E\left[\xi\right], \text{ where } E\left[\xi\right] \triangleq \int_{\mathbb{C}^n} x dP_{\xi}\left(x\right) \in \mathbb{C}^n.$$

$$Q = E\left[\left(\xi-\mu\right)\left(\xi-\mu\right)^{\dagger}\right],$$

where

$$E\left[\left(\xi-\mu\right)\left(\xi-\mu\right)^{\dagger}\right] \triangleq \int_{\mathbb{C}^n} \left(x-\mu\right)\left(x-\mu\right)^{\dagger} dP_{\xi}\left(x\right) \in \mathbb{C}^{n\times n}.$$

A complex Gaussian random vector ξ is said to be circularly symmetric if the covariance of the corresponding vector $\hat{\xi}$ has the structure

$$E\left[\left(\hat{\xi}-E\left[\hat{\xi}\right]\right)\left(\hat{\xi}-E\left[\hat{\xi}\right]\right)^T\right] = \frac{1}{2}\begin{pmatrix} \mathbb{R}e\left(Q\right) & -\mathbb{I}m(Q) \\ \mathbb{I}m(Q) & \mathbb{R}e\left(Q\right) \end{pmatrix} \tag{A.1}$$

for some Hermitian positive semidefinite matrix $Q \in \mathbb{C}^{n\times n}$. Note that the real part of a Hermitian matrix is symmetric, and the imaginary part of a Hermitian matrix is skew-symmetric. Thus the matrix appearing in (A.1) is real and symmetric. In this case $E\left[\left(\xi-E\left[\xi\right]\right)\left(\xi-E\left[\xi\right]\right)^{\dagger}\right] = Q$, and thus, a circularly symmetric complex Gaussian random vector ξ is specified by its mean and variance.

Let ξ be a circularly symmetric complex Gaussian random vector. Then the probability density function (with respect to the Radon-Nikodym derivative of the standard Lebesgue measure on $\mathbb{C}^n$) of a circularly symmetric complex Gaussian random vector with mean μ and covariance Q is derived by the following:

Due to the definition of ξ and the relationship between $\hat{\xi}$ and ξ,

$$f_{\hat{\xi}}\left(\hat{x};\hat{\mu},\hat{Q}\right)=\frac{1}{(2\pi)^{\frac{2n}{2}}\left[\det\left(\frac{1}{2}\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\hat{x}-\hat{\mu})^{\dagger}\left(\frac{1}{2}\hat{Q}\right)^{-1}(\hat{x}-\hat{\mu})\right\}$$
$$=\frac{1}{\pi^{n}\left[\det\left(\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-(\hat{x}-\hat{\mu})^{\dagger}\left(\hat{Q}\right)^{-1}(\hat{x}-\hat{\mu})\right\}.$$

The next step is to simplify the exponent part from the previous exponent function in order to finally remove the symbol "ˆ". This can be done by utilizing the following algebraic property.

Due to $C=A^{-1}$ being equivalent to $\hat{C}=\left(\hat{A}\right)^{-1}$, we have

$$\frac{1}{\pi^{n}\left[\det\left(\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-(\hat{x}-\hat{\mu})^{\dagger}\left(\hat{Q}\right)^{-1}(\hat{x}-\hat{\mu})\right\}$$
$$=\frac{1}{\pi^{n}\left[\det\left(\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-(\hat{x}-\hat{\mu})^{\dagger}\widehat{\left(Q^{-1}\right)}(\hat{x}-\hat{\mu})\right\}.$$

The next step is to remove the symbol " ˆ " in the preceding exponent function. Due to $z=x+y$ being equivalent to $\hat{z}=\hat{x}+\hat{y}$, $\mathbb{R}e\left(x^{\dagger}y\right)=\hat{x}^{\dagger}\hat{y}$ and $y=Ax$ being equivalent to $\hat{y}=\hat{A}\hat{x}$, we have

$$\frac{1}{\pi^{n}\left[\det\left(\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-(\hat{x}-\hat{\mu})^{\dagger}\widehat{\left(Q^{-1}\right)}(\hat{x}-\hat{\mu})\right\}$$
$$=\frac{1}{\pi^{n}\left[\det\left(\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-(x-\mu)^{\dagger}Q^{-1}(x-\mu)\right\}.$$

We may remove the symbol "ˆ" in the determinant: Due to $\det\left(\hat{A}\right)=|\det(A)|^{2}$, we have

$$\frac{1}{\pi^{n}\left[\det\left(\hat{Q}\right)\right]^{\frac{1}{2}}}\exp\left\{-(x-\mu)^{\dagger}Q^{-1}(x-\mu)\right\}$$
$$=\frac{1}{\pi^{n}\det(Q)}\exp\left\{-(x-\mu)^{\dagger}Q^{-1}(x-\mu)\right\}.$$

Therefore,

$$f_{\hat{\xi}}\left(\hat{x};\hat{\mu},\hat{Q}\right)=\frac{1}{\pi^{n}\det(Q)}\exp\left\{-(x-\mu)^{\dagger}Q^{-1}(x-\mu)\right\}$$
$$=\det(\pi Q)^{-1}\exp\left\{-(x-\mu)^{\dagger}Q^{-1}(x-\mu)\right\}.$$

According to the uniqueness of the Radon-Nikodym derivative and the correspondence between $\mathbb{R}^{2n}$ and $\mathbb{C}^n$, we have

$$f_\xi\left(x;\mu,Q\right)=\det\left(\pi Q\right)^{-1}\exp\left\{-\left(x-\mu\right)^\dagger Q^{-1}\left(x-\mu\right)\right\},$$

as the probability density function of ξ.

Remark A.1. Only due to the uniqueness of the Radon-Nikodym derivative, mentioned above, in measure theory, may we acquire the probability density distribution of the complex random vector.

A.2 Appendix-II: Maximum of Entropy

The channel capacity is dependent on the definition of the mutual information. At the same time the mutual information can be computed also by introducing the differential entropy and the conditional entropy. Therefore, the mutual information, the differential entropy and the conditional entropy are revisited, referring to [3] and reference therein. If familiarity of these concepts are assumed, we may skip them over to (A.2).

Definition A.1 (Mutual Information). Assume that $\xi\in\mathbb{C}^n$ and $\eta\in\mathbb{C}^n$ are two complex continuous random vectors, and $p(x)$ and $p(y)$ are the corresponding probability density functions. The mutual information $\mathbb{I}(\xi;\eta)$ between the two random vectors is defined as:

$$\mathbb{I}\left(\xi;\eta\right)\triangleq\int_{\mathbb{C}^n}\int_{\mathbb{C}^n}p\left(x\right)p\left(y|x\right)\log\frac{p\left(x\right)p\left(y|x\right)}{p\left(x\right)p\left(y\right)}dxdy,$$

where $p(y|x)$ denotes the conditional probability density function.

In information theory, the mutual information of two random variables (or vectors) is a quantity that measures the mutual dependence of the two variables (or vectors).

In information theory, the following concept of the differential entropy is measurement for the entropy of a random variable (or vector).

Definition A.2 (Differential Entropy). Assume that ξ is a complex continuous random vector, and $p(x)$ is the corresponding probability density function. Then

$$\mathbb{H}\left(\xi\right)\triangleq-\int_{\mathbb{C}^n}p\left(x\right)\log p\left(x\right)dx$$

is called the differential entropy of ξ.

In information theory, the conditional entropy quantifies the remaining entropy of a random variable (or vector) ξ given that the value of a second random variable (or vector) η is known.

Definition A.3 (Conditional Entropy). Assume that ξ and η are two complex continuous random vectors, $p(x)$ and $p(y)$ are the corresponding probability density functions and $p(x,y)$ is the corresponding joint probability density function. Then the conditional entropy of ξ for given η is defined as:

$$\mathbb{H}(\xi|\eta) \triangleq \int_{\mathbb{C}^n}\int_{\mathbb{C}^n} p(x,y)\log p(x|y)\,dxdy.$$

The following proposition offers the mathematical relationship among the mutual information, the differential entropy and the conditional entropy.

Proposition A.1. *Assume that ξ and η are two continuous random vectors. Then*

$$\mathbb{I}(\xi;\eta) = \mathbb{H}(\xi) - \mathbb{H}(\xi|\eta)$$

and

$$\mathbb{I}(\xi;\eta) = \mathbb{H}(\eta) - \mathbb{H}(\eta|\xi).$$

The differential entropy of a complex Gaussian variable (or vector) ξ with mean μ and covariance Q is derived as follows. Due to the definition of $\mathbb{H}(\xi;\mu,Q)$, we have

$$\mathbb{H}(\xi;\mu,Q) = E_\xi\left[-\log f_\xi(\xi;\mu,Q)\right], \tag{A.2}$$

where E_ξ is the expectation operator of ξ, i.e., $E_\xi[\xi] \triangleq \int_{\mathbb{C}^n} x p_\xi(x)\,dx$.

Due to the form of the probability density function of the circularly symmetric complex Gaussian random vector ξ, we have

$$E_\xi\left[-\log f_\xi(\xi;\mu,Q)\right] = \log\det(\pi Q) + E\left[(\xi-\mu)^\dagger Q^{-1}(\xi-\mu)\right].$$

Due to the definition and basic properties of the trace operator, we may write

$$\begin{aligned}\log\det(\pi Q) &+ E\left[(\xi-\mu)^\dagger Q^{-1}(\xi-\mu)\right]\\ &= \log\det(\pi Q) + E\left[\mathrm{Tr}\left((\xi-\mu)(\xi-\mu)^\dagger Q^{-1}\right)\right].\end{aligned}$$

Due to the commutative property for the product of the trace and expectation operators, we have

$$\log\det(\pi Q) + E\left[\mathrm{Tr}\left((\xi-\mu)(\xi-\mu)^{\dagger} Q^{-1}\right)\right] = \log\det(\pi Q) + \mathrm{Tr}\left(E\left[(\xi-\mu)(\xi-\mu)^{\dagger}\right] Q^{-1}\right).$$

Then the definition of the covariance of ξ implies that

$$\log\det(\pi Q) + \mathrm{Tr}\left(E\left[(\xi-\mu)(\xi-\mu)^{\dagger}\right] Q^{-1}\right) = \log\det(\pi Q) + \mathrm{Tr}\left(QQ^{-1}\right),$$

and using the definition of the logarithm function and the fact, $e \triangleq \lim_{m\to\infty}\left(1+\frac{1}{m}\right)^{m}$, we have

$$\log\det(\pi Q) + \mathrm{Tr}\left(QQ^{-1}\right) = \log\det(\pi Q) + \log e^{n}.$$

This can be simplified to

$$\log\det(\pi Q) + \log e^{n} = \log\det(\pi e Q).$$

The following proposition, which states that a circularly symmetric complex Gaussian variable (or vector) is the entropy maximizer, highlights the importance of circularly symmetric complex Gaussian vectors. Telatar [1] also claims this proposition. For proving that a circularly symmetric complex Gaussian variable (or vector) is the entropy maximizer, we can use the argument, i.e., $\log\gamma_Q(x)$ is a linear combination of the terms $x_i x_j^*$. But it is incorrect and unnecessary. In addition, the last step of deriving $\mathbb{H}(p) - \mathbb{H}(\gamma_Q) \le 0$, and $\mathbb{H}(p) - \mathbb{H}(\gamma_Q) = 0$ implying $p = \gamma_Q$ are not proved. Thus, we offer a formal and alternative proof.

Proposition A.2. *Suppose that the complex random vector $\xi \in \mathbb{C}^n$ has zero mean and ξ satisfies $E\left[\xi\xi^{\dagger}\right] = Q$, i.e., $E\left[\xi_i\xi_j^{\dagger}\right] = Q_{i,j}, 1 \le i, j \le n$. Then the entropy of ξ satisfies $\mathbb{H}(f(\xi;\mu,Q)) \le \log\det(\pi e Q)$, with equality if and only if ξ is a circularly symmetric complex Gaussian random variable (or vector).*

The following two important facts are needed to complete our proof. *The first important fact is:*

$$E_\eta\left[\log p_\eta(\eta)\right] = E_\xi\left[\log p_\eta(\xi)\right].$$

The second one is a simple but crucial inequality in our proof. *The second important fact is:*

$$\log x \le x - 1, \forall x > 0;\ \log x = x - 1,\ \text{iff } x = 1.$$

The first one needs a proof that is given as follows.

Lemma A.1. *If assumptions are the same as those of Proposition A.2, the random vector ξ and η satisfy the assumptions and η is a circularly symmetric complex Gaussian random vector, then*

$$E_\eta\left[\log p_\eta\left(\eta\right)\right] = E_\xi\left[\log p_\eta\left(\xi\right)\right].$$

Proof (The Proof of Lemma A.1). Due to the definitions of η and (A.2), which qualify the relationship expression between the differential entropy and the mean, we have

$$E_\eta\left[\log p_\eta\left(\eta\right)\right] = \int_{\mathbb{C}^n}\left[-\log\det\left(\pi Q\right) - \left(x-\mu\right)^\dagger Q^{-1}\left(x-\mu\right)\right] p_\eta\left(x\right) dx.$$

Due to the linearity property of integration, we have

$$\int_{\mathbb{C}^n}\left[-\log\det\left(\pi Q\right) - \left(x-\mu\right)^\dagger Q^{-1}\left(x-\mu\right)\right] p_\eta\left(x\right) dx$$
$$= -\log\det\left(\pi Q\right) - \int_{\mathbb{C}^n}\left(x-\mu\right)^\dagger Q^{-1}\left(x-\mu\right) p_\eta\left(x\right) dx.$$

Due to the circular invariance property of the trace, $\mathrm{Tr}\left(ABC\right) = \mathrm{Tr}\left(BCA\right)$, we have

$$-\log\det\left(\pi Q\right) - \int_{\mathbb{C}^n}\left(x-\mu\right)^\dagger Q^{-1}\left(x-\mu\right) p_\eta\left(x\right) dx$$
$$= -\log\det\left(\pi Q\right) - \int_{\mathbb{C}^n}\mathrm{Tr}\left(Q^{-1}\left(x-\mu\right)\left(x-\mu\right)^\dagger\right) p_\eta\left(x\right) dx.$$

Because the order of the trace and integration can be interchanged, one has

$$-\log\det\left(\pi Q\right) - \int_{\mathbb{C}^n}\mathrm{Tr}\left(Q^{-1}\left(x-\mu\right)\left(x-\mu\right)^\dagger\right) p_\eta\left(x\right) dx$$
$$= -\log\det\left(\pi Q\right) - \mathrm{Tr}\left(\int_{\mathbb{C}^n} Q^{-1}\left(x-\mu\right)\left(x-\mu\right)^\dagger p_\eta\left(x\right) dx\right).$$

Due to the linearity property of the integration, we have

$$-\log\det\left(\pi Q\right) - \mathrm{Tr}\left(\int_{\mathbb{C}^n} Q^{-1}\left(x-\mu\right)\left(x-\mu\right)^\dagger p_\eta\left(x\right) dx\right)$$
$$= -\log\det\left(\pi Q\right) - \mathrm{Tr}\left(Q^{-1}\int_{\mathbb{C}^n}\left(x-\mu\right)\left(x-\mu\right)^\dagger p_\eta\left(x\right) dx\right).$$

The assumption that the variances of ξ and η are the same implies

$$Q = \int_{\mathbb{C}^n} (x-\mu)(x-\mu)^\dagger p_\eta(x)\,dx = \int_{\mathbb{C}^n} (x-\mu)(x-\mu)^\dagger p_\xi(x)\,dx,$$

$$-\log\det(\pi Q) - \mathrm{Tr}\left(Q^{-1}\int_{\mathbb{C}^n} (x-\mu)(x-\mu)^\dagger p_\eta(x)\,dx\right)$$
$$= \int_{\mathbb{C}^n} -\log\det(\pi Q)\, p_\xi(x)\,dx + \mathrm{Tr}\left(Q^{-1}\int_{\mathbb{C}^n} -(x-\mu)(x-\mu)^\dagger p_\xi(x)\,dx\right).$$

Because of the basic property, $\log(ab) = \log(a) + \log(b)$, of the logarithm, it is obtained that

$$\int_{\mathbb{C}^n} -\log\det(\pi Q)\, p_\xi(x)\,dx + \mathrm{Tr}\left(Q^{-1}\int_{\mathbb{C}^n} -(x-\mu)(x-\mu)^\dagger p_\xi(x)\,dx\right)$$
$$= \int_{\mathbb{C}^n} \log\left(\det(\pi Q)^{-1}\exp\left\{\mathrm{Tr}\left(-Q^{-1}(x-\mu)(x-\mu)^\dagger\right)\right\}\right) p_\xi(x)\,dx.$$

The circular invariance property, $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$, of the trace implies

$$\int_{\mathbb{C}^n} \log\left(\det(\pi Q)^{-1}\exp\left\{\mathrm{Tr}\left(-Q^{-1}(x-\mu)(x-\mu)^\dagger\right)\right\}\right) p_\xi(x)\,dx$$
$$= \int_{\mathbb{C}^n} \log\left(\det(\pi Q)^{-1}\exp\left\{-(x-\mu)^\dagger Q^{-1}(x-\mu)\right\}\right) p_\xi(x)\,dx.$$

Finally, due to the definition of $E_\xi[\log p_\eta(\xi)]$, i.e.,

$$\int_{\mathbb{C}^n} \log\left(\det(\pi Q)^{-1}\exp\left\{-(x-\mu)^\dagger Q^{-1}(x-\mu)\right\}\right) p_\xi(x)\,dx = E_\xi[\log p_\eta(\xi)],$$

we have $E_\eta[\log p_\eta(\eta)] = E_\xi[\log p_\eta(\xi)]$.

The proof of Proposition A.2 is offered as follows.

Proof. Let $p_\xi : \mathbb{C}^n \longrightarrow \mathbb{R}$ be the probability density function of ξ. According to the assumption of the random vector ξ,

$$E\left(\xi\xi^\dagger\right) = \int_{\mathbb{C}^n} xx^\dagger p_\xi(x)\,dx = Q.$$

Let η be a circularly symmetric complex Gaussian variable (or vector) with zero mean and variance $E\left(\eta\eta^\dagger\right) = Q$. Let the probability density function of η be $p_\eta(x)$.

The definitions of $\mathbb{H}(\xi)$ and $\mathbb{H}(\eta)$ imply

$$\mathbb{H}(\xi) - \mathbb{H}(\eta) = -E_\xi\left[\log p_\xi(\xi)\right] + E_\eta[\log p_\eta(\eta)]. \tag{A.3}$$

The first important fact holding is followed by

$$-E_{\xi}\left[\log p_{\xi}(\xi)\right]+E_{\eta}\left[\log p_{\eta}(\eta)\right]=-E_{\xi}\left[\log p_{\xi}(\xi)\right]+E_{\xi}\left[\log p_{\eta}(\xi)\right].$$

Because of the linearity property of the expectation,

$$-E_{\xi}\left[\log p_{\xi}(\xi)\right]+E_{\xi}\left[\log p_{\eta}(\xi)\right]=E_{\xi}\left[\log\frac{p_{\eta}(\xi)}{p_{\xi}(\xi)}\right].$$

Due to the second important fact holding and basic properties of Lebesgue integration, it is to see

$$E_{\xi}\left[\log\frac{p_{\eta}(\xi)}{p_{\xi}(\xi)}\right]\le E_{\xi}\left[\frac{p_{\eta}(\xi)}{p_{\xi}(\xi)}-1\right]. \tag{A.4}$$

The definition of the expectation implies

$$E_{\xi}\left[\frac{p_{\eta}(\xi)}{p_{\xi}(\xi)}-1\right]=E_{\xi}\left[\frac{p_{\eta}(\xi)}{p_{\xi}(\xi)}\right]-E_{\xi}[1]=1-1=0.$$

Hence, according to (A.3), $\mathbb{H}(\xi)-\mathbb{H}(\eta)\le 0$.

Therefore, taking note of the second important fact, the entropy of ξ satisfies

$$\mathbb{H}(\xi)=-E_{\xi}\left[\log p_{\xi}(\xi)\right]\le\log\det(\pi e Q),$$

with equality if and only if ξ is a circularly symmetric complex Gaussian random variable (or vector) under the given mean and variance.

Remark A.2. Equation (A.4) is explained as follows. First, define $u\log u|_{u=0}\triangleq\lim_{u\downarrow 0}u\log u=0$ and $\frac{u}{u}|_{u=0}\triangleq\lim_{u\downarrow 0}\frac{u}{u}=1$. Second, (A.4) holds because

$$\begin{aligned}
E_{\xi}\left[\log\frac{p_{\eta}(\xi)}{p_{\xi}(\xi)}\right] &= \int_{\mathbb{C}^n}\log\left(\frac{p_{\eta}(x)}{p_{\xi}(x)}\right)p_{\xi}(x)\,dx\\
&= \left(\int_{\{x|p_{\xi}(x)=0\}}+\int_{\{x|p_{\xi}(x)>0\}}\right)\log\left(\frac{p_{\eta}(x)}{p_{\xi}(x)}\right)p_{\xi}(x)\,dx\\
&= \int_{\{x|p_{\xi}(x)=0\}}0dx+\int_{\{x|p_{\xi}(x)>0\}}\log\left(\frac{p_{\eta}(x)}{p_{\xi}(x)}\right)p_{\xi}(x)\,dx\\
&\le \int_{\{x|p_{\xi}(x)>0\}}\left(\frac{p_{\eta}(x)}{p_{\xi}(x)}-1\right)p_{\xi}(x)\,dx
\end{aligned}$$

$$\leq \int_{\mathbb{C}^n} \frac{p_\eta(x)}{p_\xi(x)} p_\xi(x)\,dx - \left(\int_{\{x|p_\xi(x)=0\}} + \int_{\{x|p_\xi(x)>0\}} \right) p_\xi(x)\,dx$$

$$= \int_{\mathbb{C}^n} \frac{p_\eta(x)}{p_\xi(x)} p_\xi(x)\,dx - \int_{\mathbb{C}^n} p_\xi(x)\,dx$$

$$= E_\xi \left[\frac{p_\eta(\xi)}{p_\xi(\xi)} \right] - E_\xi[1]$$

$$= E_\xi \left[\frac{p_\eta(\xi)}{p_\xi(\xi)} - 1 \right].$$

Using the definition of the mutual information, we have that

$$C(H,P) = \max_{p_x} \{ \mathbb{H}(y) - \mathbb{H}(y|x) \,| S \succeq 0, \mathrm{Tr}(S) \leq P \},$$

where the model (3.1) implies

$$\max_{p_x} \{ \mathbb{H}(y) - \mathbb{H}(y|x) \,| S \succeq 0, \mathrm{Tr}(S) \leq P \}$$
$$= \max_{p_x} \{ \mathbb{H}(y) - \mathbb{H}(z) \,| S \succeq 0, \mathrm{Tr}(S) \leq P \}.$$

The assumptions of z in model (3.1) also imply

$$\max_{p_x} \{ \mathbb{H}(y) - \mathbb{H}(z) \,| S \succeq 0, \mathrm{Tr}(S) \leq P \}$$
$$= \max_{p_x} \{ \mathbb{H}(y) \,| S \succeq 0, \mathrm{Tr}(S) \leq P \} - \mathbb{H}(z).$$

Note that we may assume that x satisfies $E\left(x^\dagger x\right) \leq P$ and is a zero mean random vector. Furthermore for such an x, if x is a zero mean random vector with covariance $E\left(xx^\dagger\right) = S$, then y is a zero mean random vector with covariance $E\left(yy^\dagger\right) = HSH^\dagger + I_r$, which results from the form of model (3.1) and the linearity of the expectation operation, and by Proposition A.2 among such y vectors the entropy is the largest when y is a circularly symmetric complex Gaussian random vector, which is the case when x is a circularly symmetric complex Gaussian random vector by the two facts at the end part of last section. Thus, we can further restrict our attention to the circularly symmetric complex Gaussian random vector x. In this case the mutual information is given by $\log\left(\det\left(I_r + HSH^\dagger\right)\right)$.

The two facts are claimed as follows. A linear transformation of a circularly symmetric complex Gaussian random vector is a circularly symmetric complex Gaussian random vector. The set of circularly symmetric complex Gaussian random vectors is closed for addition. They are used for calculating the channel capacity in the following section.

A.3 Appendix-III: Proofs of the Lemmas in Sect. 3.1.3

Lemma A.2. *For the channel H, there is a unitary matrix U such that $U^\dagger H^\dagger HU = diag(\lambda_1, \cdots, \lambda_t)$ (the diagonal matrix) and*

$$\max\left\{\log\left(\det\left(I_r + HSH^\dagger\right)\right) | S \succeq 0, Tr(S) \le P\right\} = \max\{\log(\det(I_t + diag(\lambda_1, \cdots, \lambda_t)S)) | S \succeq 0, Tr(S) \le P\}$$

and $U^\dagger S_l U = S_r$, where S_l and S_r are two optimal solutions of the two optimization problems mentioned above, respectively.

Proof. According to the matrix theory, it is easily known that there is a unitary matrix U such that

$$U^\dagger H^\dagger HU = \mathrm{diag}(\lambda_1, \cdots, \lambda_t) \tag{A.5}$$

(the diagonal matrix) and the two maximum points exist due to the compactness of the two constraints. Let

$$S_l \in \arg\max\left\{\log\left(\det\left(I_r + HSH^\dagger\right)\right) | S \succeq 0, \mathrm{Tr}(S) \le P\right\}.$$

Because U is a unitary matrix and $\det(I + AB) = \det(I + BA)$ with appropriate dimensions of the matrices, we have

$$\log\left(\det\left(I_r + HS_lH^\dagger\right)\right) = \log\left(\det\left(I_t + U^\dagger H^\dagger HUU^\dagger S_l U\right)\right).$$

Due to (A.5),

$$\log\left(\det\left(I_t + U^\dagger H^\dagger HUU^\dagger S_l U\right)\right) = \log\left(\det\left(I_t + \mathrm{diag}(\lambda_1, \cdots, \lambda_t) U^\dagger S_l U\right)\right).$$

Since $S_l \succeq 0$ and $\mathrm{Tr}(S_l) \le P$,

$$\log\left(\det\left(I_t + \mathrm{diag}(\lambda_1, \cdots, \lambda_t) U^\dagger S_l U\right)\right) \le \max\left\{\log\left(\det\left(I_t + \mathrm{diag}(\lambda_1, \cdots, \lambda_t) U^\dagger SU\right)\right) | U^\dagger SU \succeq 0, \mathrm{Tr}\left(U^\dagger SU\right) \le P\right\}.$$

As the unitary similarity transformation keeps the semidefinite positiveness and trace,

$$\max\left\{\log\left(\det\left(I_t + \mathrm{diag}(\lambda_1, \cdots, \lambda_t) U^\dagger SU\right)\right) | U^\dagger SU \succeq 0, \mathrm{Tr}\left(U^\dagger SU\right) \le P\right\} \le \max\{\log(\det(I_t + \mathrm{diag}(\lambda_1, \cdots, \lambda_t) S)) | S \succeq 0, \mathrm{Tr}(S) \le P\}.$$

Hence,

$$\max\left\{\log\left(\det\left(I_r+HSH^\dagger\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}\leq$$
$$\max\left\{\log\left(\det\left(I_t+\mathrm{diag}\left(\lambda_1,\cdots,\lambda_t\right)S\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}.$$

On the other hand,

$$\forall S_r, S_r\in\arg\max\left\{\log\left(\det\left(I_t+\mathrm{diag}\left(\lambda_1,\cdots,\lambda_t\right)S\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}.$$

Because the definition of matrix Λ,

$$\log\left(\det\left(I_t+\mathrm{diag}\left(\lambda_1,\cdots,\lambda_t\right)S_r\right)\right)\leq\log\left(\det\left(I_t+U^\dagger H^\dagger HUS_r\right)\right).$$

Due to $\det(I+AB)=\det(I+BA)$,

$$\log\left(\det\left(I_t+U^\dagger H^\dagger HUS_r\right)\right)\leq\log\left(\det\left(I_r+HUS_rU^\dagger H^\dagger\right)\right).$$

As the unitary similarity transformation keeps the semidefinite positiveness and trace, we get

$$\log\left(\det\left(I_r+HUS_rU^\dagger H^\dagger\right)\right)$$
$$\leq\max\left\{\log\left(\det\left(I_r+HUS_rU^\dagger H^\dagger\right)\right)|US_rU^\dagger\succeq 0,\mathrm{Tr}\left(US_rU^\dagger\right)\leq P\right\}.$$

The same reason implies

$$\max\left\{\log\left(\det\left(I_r+HUS_rU^\dagger H^\dagger\right)\right)|US_rU^\dagger\succeq 0,\mathrm{Tr}\left(US_rU^\dagger\right)\leq P\right\}$$
$$\leq\max\left\{\log\left(\det\left(I_r+HSH^\dagger\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}.$$

Thus,

$$\max\left\{\log\left(\det\left(I_t+\mathrm{diag}\left(\lambda_1,\cdots,\lambda_t\right)S\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}$$
$$\leq\max\left\{\log\left(\det\left(I_r+HSH^\dagger\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}.$$

Therefore,

$$\max\left\{\log\left(\det\left(I_r+HSH^\dagger\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}$$
$$=\max\left\{\log\left(\det\left(I_t+\mathrm{diag}\left(\lambda_1,\cdots,\lambda_t\right)S\right)\right)|S\succeq 0,\mathrm{Tr}\left(S\right)\leq P\right\}$$

and further $U^\dagger S_l U=S_r$, where S_l and S_r are two optimal solutions of the two optimization problems respectively.

Lemma A.3. *For the channel H, there is a unitary matrix U such that $U^\dagger H^\dagger HU = \Lambda$ (the diagonal matrix) and*

$$\max\left\{\log\left(\det\left(I_r + HSH^\dagger\right)\right) | S \succeq 0, Tr(S) \leq P\right\}$$
$$= \max\left\{\log\left(\det\left(I_t + \Lambda^{\frac{1}{2}} S \Lambda^{\frac{1}{2}}\right)\right) | S \succeq 0, Tr(S) \leq P\right\},$$

and $U^\dagger S_l U = S_r$, where S_l and S_r are two optimal solutions of the two optimization problems mentioned above, respectively.

Proof. According to **Lemma A.2**, we have

$$\max\left\{\log\left(\det\left(I_r + HSH^\dagger\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}$$
$$= \max\left\{\log\left(\det\left(I_t + \mathrm{diag}\left(\lambda_1, \cdots, \lambda_t\right) S\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}.$$

According to the definition of the matrix Λ and (A.5), we have

$$\max\left\{\log\left(\det\left(I_t + \mathrm{diag}\left(\lambda_1, \cdots, \lambda_t\right) S\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}$$
$$= \max\left\{\log\left(\det\left(I_t + \Lambda S\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}.$$

Due to the definition of the square root for the matrix Λ, we get

$$\max\left\{\log\left(\det\left(I_t + \Lambda S\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}$$
$$= \max\left\{\log\left(\det\left(I_t + \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} S\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}.$$

For the reason that $\det(I + AB) = \det(I + BA)$, we have

$$\max\left\{\log\left(\det\left(I_t + \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} S\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}$$
$$= \max\left\{\log\left(\det\left(I_t + \Lambda^{\frac{1}{2}} S \Lambda^{\frac{1}{2}}\right)\right) | S \succeq 0, \mathrm{Tr}(S) \leq P\right\}.$$

References

1. E. Telatar, "Capacity of multi-antenna Gaussian channels," European Transactions on Telecommunications, vol. 10, pp. 585–596, 1999.
2. J. L. Doob, Measure Theory (Graduate Texts in Mathematics), Springer-Verlag, New York, 1994.
3. E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj and H. V. Poor, MIMO Wireless Communications, Cambridge University Press, Cambridge, 2007.