Equality in the matrix entropy-power inequality and blind separation of real and complex sources

Olivier Rioul, Ram Zamir

Equality in the Matrix Entropy-Power Inequality and Blind Separation of Real and Complex sources

Olivier Rioul
LTCI, Télécom Paris
Institut Polytechnique de Paris, 75013, Paris, France
Email: olivier.rioul@telecom-paristech.fr

Ram Zamir
EE - Systems Department
Tel Aviv University, Tel Aviv, Israel
Email: zamir@eng.tau.ac.il

***Abstract*—The matrix version of the entropy-power inequality for real or complex coefficients and variables is proved using a transportation argument that easily settles the equality case. An application to blind source extraction is given.**

I. Introduction

Consider random variables with densities that are continuous and positive inside their support interval, with zero mean and finite differential entropies. The entropy power inequality (EPI) was stated by Shannon [1] in 1948 and is well known to be equivalent to the following minimum entropy inequality [2]–[4]:

$$h(a_1X_1 + a_2X_2) \geq h(a_1X_1^* + a_2X_2^*) \tag{1}$$

for any real numbers a_1, a_2 and any independent real random variables X_1, X_2, where X_1^*, X_2^* are independent normal random variables having the same entropies as X_1, X_2:

$$h(X_1^*) = h(X_1) \qquad h(X_2^*) = h(X_2). \tag{2}$$

Equality holds in (1) if and only if either $a_1a_2 = 0$ or X_1, X_2 are normal. Recently, a normal transport argument was used in [5] to provide a simple proof of Shannon's EPI, including the necessary and sufficient condition for equality.

Shannon's EPI was generalized to a matrix version [6], [7]:

$$h(\mathbf{A}X) \geq h(\mathbf{A}X^*) \tag{3}$$

for any $m \times n$ matrix $\mathbf{A}$ and any random (column) vector $X = (X_1, X_2, \ldots, X_n)^t$ of independent components X_i, where $X^* = (X_1^*, X_2^*, \ldots, X_n^*)^t$ is a normal vector with independent components X_i^* of the same entropies:

$$h(X_i^*) = h(X_i) \qquad (i = 1, \ldots, n). \tag{4}$$

Available proofs of (3) are either by double induction on (m, n) [6] or by integration over a path of Gaussian perturbation of the corresponding inequality for Fisher's information using de Bruijn's identity [7] or via the I-MMSE relation [8]. A necessary and sufficient condition for equality in (3) has not been settled so far, however, by the previous methods. Such a condition is important in applications such as blind source separation (BSS) based on minimum entropy [9]. Also, BSS may involve real or complex signals [10] and minimum entropy methods for complex sources would require the extension of EPIs to complex-valued variables and coefficients.

In this paper, we adapt the proof of [5] to the matrix case and derive (3) with a normal transport argument. This allows us to easily settle the equality case: We define the notion of "recoverability" and show that equality holds in (3) if all unrecoverable components of X present in $\mathbf{A}X$ are normal. We then extend the proofs to complex-valued $\mathbf{A}$ and X. As an application, we derive the appropriate contrast functions for partial BSS (a.k.a. blind source extraction) where m out of n independent sources are to be extracted.

II. A simple proof of the matrix EPI by transport

We extend the proof in [5] to the matrix EPI, based on the same ingredients: (a) a transportation argument from normal variables, that takes the form of a simple change of variables; (b) a rotation performed on i.i.d. normal variables, which preserves the i.i.d. property; (c) concavity of the logarithm, appropriately generalized to the matrix case. The proof breaks into several elementary steps:

A. Reduce to full rank $m < n$

If the rank of $\mathbf{A}$ is $< m$ then some rows are linearly dependent, there is a deterministic relation between some components of $\mathbf{A}X$ and $\mathbf{A}X^*$ and equality $h(\mathbf{A}X) = h(\mathbf{A}X^*) = -\infty$ holds trivially. Thus we can assume that $\mathbf{A}$ is of full rank $m \leq n$. If $\mathbf{A}$ has rank $m = n$ then $\mathbf{A}$ is invertible and by the change of variable formula in the entropy [1, § 20.9], $h(\mathbf{A}X) = h(X) + \log|\mathbf{A}| = h(X^*) + \log|\mathbf{A}| = h(\mathbf{A}X^*)$ where $|\mathbf{A}|$ denotes the absolute value of the determinant of $\mathbf{A}$. Therefore, one may always assume that $\mathbf{A}$ has full rank $m < n$.

B. Reduce to equal individual entropies

Without loss of generality, one may assume that the components of X have equal entropies. For if it were not the case, then by the scaling property of entropy [1, § 20.9], one can find non zero coefficients δ_j (e.g., $\delta_j = \exp h(X_j)$) such that all $X_j' = X_j/\delta_j$ have equal entropies. Then applying (3) to $X' = (X_1', \ldots, X_n')^t$ and matrix $\mathbf{A}\mathbf{\Delta}$ where $\mathbf{\Delta}$ is a diagonal matrix with diagonal elements δ_j, gives the desired EPI.

Notice that with the additional constraint that the X_j have equal entropies, we have $h(X_1^*) = h(X_2^*) = \cdots = h(X_n^*) = h(X_1) = h(X_2) = \cdots = h(X_n)$: The independent zero-mean normal variables X_j^* also have equal entropies, and are, therefore, independent and identically distributed (i.i.d.).

C. Reduce to orthonormal rows

Without loss of generality, one may assume that the rows of $\mathbf{A}$ are orthonormal. For if it were not the case, one can orthonormalize the rows by a Gram-Schmidt process. This amounts to multiplying $\mathbf{A}$ on the left by an lower-triangular invertible matrix $\mathbf{L}$. Thus, one can apply (3) for matrix $\mathbf{A}' = \mathbf{L}\mathbf{A}$. Again by the change of variable in the entropy [1, § 20.9], $h(\mathbf{A}'X) = h(\mathbf{A}X) + \log|\mathbf{L}|$ and $h(\mathbf{A}'X^*) = h(\mathbf{A}X^*) + \log|\mathbf{L}|$. The terms $\log|\mathbf{L}|$ cancel to give the desired EPI. Thus we are led to prove (3) for an $m \times n$ matrix $\mathbf{A}$ with orthonormal rows ($\mathbf{A}\mathbf{A}^t = \mathbf{I}_m$, the $m \times m$ identity matrix).

D. Complete the orthogonal matrix

Extend $\mathbf{A}$ by adding $n - m$ orthonormal rows of a complementary matrix $\mathbf{A}'$ such that $\left(\frac{\mathbf{A}}{\mathbf{A}'}\right)$ is an $n \times n$ orthogonal matrix, and define the Gaussian vector $\left(\frac{\widetilde{X}}{\widetilde{X}'}\right)$ as

$$\left(\frac{\widetilde{X}}{\widetilde{X}'}\right) = \left(\frac{\mathbf{A}}{\mathbf{A}'}\right) X^*. \tag{5}$$

Since the components of X^* are i.i.d. normal and $\left(\frac{\mathbf{A}}{\mathbf{A}'}\right)$ is orthogonal, the components of $\left(\frac{\widetilde{X}}{\widetilde{X}'}\right)$ are also i.i.d. normal. In particular the subvectors $\widetilde{X}$ and $\widetilde{X}'$ are independent. The inverse transformation is the transpose:

$$X^* = \left(\mathbf{A}^t \;\middle|\; \mathbf{A}'^{\,t}\right) \left(\frac{\widetilde{X}}{\widetilde{X}'}\right) = \mathbf{A}^t\widetilde{X} + \mathbf{A}'^{\,t}\widetilde{X}'. \tag{6}$$

E. Apply the normal transportation

Lemma 1 (Normal Transportation [5], [11]): *Let $X^* \in \mathbb{R}$ be a scalar normal random variable. For any continuous density f, there exists a differentiable transformation $T\colon \mathbb{R} \to \mathbb{R}$ with positive derivative $T' > 0$ such that $X = T(X^*)$ has density f.*

From Lemma 1, we can assume that the components of $X = (X_1, X_2, \dots, X_n)^t$ and $X^* = (X_1^*, X_2^*, \dots, X_n^*)^t$ are such that $X_j = T_j(X_j^*)$ for all $j = 1, 2, \dots, n$, where the T_j's are transformations with positive derivatives $T_j' > 0$. For ease of notation define

$$T(X^*) = \big(T_1(X_1^*), T_2(X_2^*), \dots, T_n(X_n^*)\big)^t \tag{7}$$

Thus $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a transformation whose Jacobian matrix is diagonal with positive diagonal elements:

$$T'(X^*) = \operatorname{diag}\big(T_1'(X_1^*), \dots, T_n'(X_n^*)\big). \tag{8}$$

Now (3) can be written in terms of the normal variables only:

$$h\big(\mathbf{A}T(X^*)\big) \geq h(\mathbf{A}X^*) \tag{9}$$

and by (6) it can also be written in term of the tilde normal variables:

$$h\big(\mathbf{A}T(\mathbf{A}^t\widetilde{X} + \mathbf{A}'^{\,t}\widetilde{X}')\big) \geq h(\widetilde{X}). \tag{10}$$

F. Conditioning on the complementary variables

Since conditioning reduces entropy [1, § 20.4],

$$h\big(\mathbf{A}T(\mathbf{A}^t\widetilde{X}+\mathbf{A}'^{\,t}\widetilde{X}')\big) \geq h\big(\mathbf{A}T(\mathbf{A}^t\widetilde{X}+\mathbf{A}'^{\,t}\widetilde{X}') \mid \widetilde{X}'\big). \tag{11}$$

G. Make the change of variable

By the change of variable formula in the entropy [1, § 20.8], $h(X_j) = h(T_j(X_j^*)) = h(X_j^*) + \mathbb{E}\log T_j'(X_j^*)$ and, therefore, by (4),

$$\mathbb{E}\log T_j'(X_j^*) = 0 \qquad (j = 1, 2, \dots, n). \tag{12}$$

By the change of variable formula (vector case) [1, § 20.8] in the conditional entropy in the r.h.s. of (11),

$$\begin{aligned} &h\big(\mathbf{A}T(\mathbf{A}^t\widetilde{X} + \mathbf{A}'^{\,t}\widetilde{X}') \mid \widetilde{X}'\big) \\ &\quad = h(\widetilde{X} \mid \widetilde{X}') + \mathbb{E}\log|\mathbf{A}T'(\mathbf{A}^t\widetilde{X} + \mathbf{A}'^{\,t}\widetilde{X}')\mathbf{A}^t| \quad (13)\\ &\quad = h(\widetilde{X}) + \mathbb{E}\log|\mathbf{A}T'(X^*)\mathbf{A}^t| \quad (14)\end{aligned}$$

where we have used that $\widetilde{X}$ and $\widetilde{X}'$ are independent.

H. Apply the concavity of the logarithm

The following lemma was stated in [7] as a consequence of (3). A direct proof was given in [8], and is simplified here.

Lemma 2: *For any $m \times n$ matrix $\mathbf{A}$ with orthonormal rows and any diagonal matrix $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ with positive diagonal elements $\lambda_j > 0$,*

$$\log|\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t| \geq \operatorname{tr}(\mathbf{A}[\log\mathbf{\Lambda}]\mathbf{A}^t) \tag{15}$$

where $\log\mathbf{\Lambda} = \operatorname{diag}(\log\lambda_1, \dots, \log\lambda_n)$ and $\operatorname{tr}(\cdot)$ denotes the trace.

Equality holds e.g. when the λ_j's are equal. The precise equality case will appear elsewhere.

Proof: It is easily checked that $\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t$ is positive definite and that both sides of (15) do not change if we replace $\mathbf{A}$ by $\mathbf{U}\mathbf{A}$ where $\mathbf{U}$ is any $m \times m$ orthogonal matrix. Choose $\mathbf{U}$ as an orthogonal eigenvector matrix of $\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t$, so that $\mathbf{U}\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t\mathbf{U}^t$ is diagonal with positive diagonal elements and $\mathbf{U}\mathbf{A}$ still has orthonormal rows.

Thus, substituting $\mathbf{U}\mathbf{A}$ for $\mathbf{A}$ we may always assume that $\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t$ is diagonal with diagonal entries equal to $\sum_{j=1}^n A_{ij}^2\lambda_j$ for $i = 1, 2, \dots, m$, where $A_{i,j}$ denotes the entries of $\mathbf{A}$. Then

$$\begin{aligned} \log|\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t| &= \sum_{i=1}^m \log\sum_{j=1}^n A_{ij}^2\lambda_j \quad (16)\\ &\geq \sum_{i=1}^m\sum_{j=1}^n A_{ij}^2\log\lambda_j \quad (17)\\ &= \operatorname{tr}(\mathbf{A}[\log\mathbf{\Lambda}]\mathbf{A}^t). \quad (18)\end{aligned}$$

where (17) follows from Jensen's inequality and the concavity of the logarithm, since $\mathbf{A}$ has orthonormal rows. □

From Lemma 2 and (12) we obtain

$$\begin{aligned} \mathbb{E}\log|\mathbf{A}T'(X^*)\mathbf{A}^t| &\geq \mathbb{E}\operatorname{tr}(\mathbf{A}[\log T'(X^*)]\mathbf{A}^t) \quad (19)\\ &= \operatorname{tr}(\mathbf{A}\mathbb{E}[\log T'(X^*)]\mathbf{A}^t) = 0. \quad (20)\end{aligned}$$

Combining this with (11)–(14) proves (10) and the desired matrix EPI (3).

III. The Equality Case

To settle the equality case in (3), from the remarks in § II-A we may already assume that $\mathbf{A}$ has full rank $m < n$.

Definition 1: *A component X_j of X is*

- present *in $\mathbf{A}X$ if $\mathbf{A}X$ depends on X_j;*
- recoverable *from $\mathbf{A}X$ if there exists a row vector b such that $b\cdot(\mathbf{A}X) = X_j$.*

Remark 1: Since the considered variables are not deterministic, Definition 1 depends only on the matrix $\mathbf{A}$: X_j is present in $\mathbf{A}X$ if and only if the jth column of $\mathbf{A}$ is not zero; and X_j is recoverable from $\mathbf{A}X$ if and only if there exists b such that $b\mathbf{A} = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the jth position. A recoverable component is necessarily present.

Remark 2: Without loss of generality we always omit the components that are not present in $\mathbf{A}X$ and their associated zero columns of $\mathbf{A}$ without affecting the entropy $h(\mathbf{A}X)$.

Remark 3: Definition 1 is also invariant by left multiplication of $\mathbf{A}$ by any $m \times m$ invertible matrix $\mathbf{B}$: if the jth column of $\mathbf{A}$ is zero, so is the jth column of $\mathbf{BA}$; and $b\mathbf{A} = (0, \dots, 0, 1, 0, \dots, 0)$ implies $(b\mathbf{B}^{-1})(\mathbf{BA}) = (0, \dots, 0, 1, 0, \dots, 0)$.

The following property was used in [12, Appendix] for deriving a sufficient condition for equality in a matrix form of the Brunn–Minkowski inequality, which is the analog of the EPI for Rényi entropies of order zero [3].

Lemma 3: *Reordering the components of X if necessary so that the first r components are recoverable and the last $n-r$ components are unrecoverable, we may always put $\mathbf{A}$ in the canonical form*

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_u \end{array}\right) \tag{21}$$

where $\mathbf{A}_u$ is an $(m-r) \times (n-r)$ matrix. The number r of recoverable components is the maximum *number such that $\mathbf{A}$ can be put in the form* (21) *by left multiplication by an invertible matrix.*

Proof: Write $X = (X_r \mid X_u)^t$ where X_r has recoverable components and X_u has unrecoverable ones. By Definition 1 (recoverability) there exists a $r \times m$ matrix $\mathbf{B}_r$ such that $\mathbf{B}_r\mathbf{A} = (\mathbf{I}_r \mid 0)$. Since $\mathbf{B}_r$ must have rank r, this shows in particular that $r \le m$: no more than m components can be recovered from the m linear mixtures. We can use $m-r$ additional row operations so that $\left(\begin{smallmatrix}\mathbf{B}_r \\ \mathbf{B}_u\end{smallmatrix}\right)\mathbf{A} = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_u \end{array}\right)$ is of the desired form. Since $\mathbf{B} = \left(\begin{smallmatrix}\mathbf{B}_r \\ \mathbf{B}_u\end{smallmatrix}\right)$ is an $m \times m$ invertible matrix, by the change of variable formula in the entropy [1, §20.9], $h(\mathbf{BA}X) = h(\mathbf{A}X) + \log|\mathbf{B}|$ and $h(\mathbf{BA}X^*) = h(\mathbf{A}X^*) + \log|\mathbf{B}|$. Therefore, the matrix EPI (3) is equivalent to the one obtained by substituting $\mathbf{BA} = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_u \end{array}\right)$ for $\mathbf{A}$. Clearly, r is maximum in this expression since otherwise one could recover more than r components, hence transfer some of the components from the $\mathbf{A}_u$ block to the $\mathbf{I}_r$ block. □

We can now settle the equality case in (3).

Theorem 1: *Equality holds in* (3) *if and only if all unrecoverable components present in $\mathbf{A}X$ are normal.*

Proof: Write $X = (X_r \mid X_u)^t$ as in the proof of Lemma 3 and accordingly write $X^* = (X_r^* \mid X_u^*)^t$. If $\mathbf{A}$ is in canonical form (21), then (3) reads

$$h(X_r) + h(\mathbf{A}_u X_u) \ge h(X_r^*) + h(\mathbf{A}_u X_u^*). \tag{22}$$

where $h(X_r) = \sum_{j=1}^r h(X_j) = \sum_{j=1}^r h(X_j^*) = h(X_r^*)$. The announced condition is, therefore, sufficient: if X_u is normal with (zero-mean) components satisfying (4), then X_u is identically distributed as X_u^* and $h(\mathbf{A}_u X_u) = h(\mathbf{A}_u X_u^*)$.

Conversely, suppose that (3) is an equality with $\mathbf{A}$ as in (21). From § II C, we may assume (applying row operations of a Gram-Schmidt process if necessary) that $\mathbf{A}$ has orthonormal rows in (21), that is, $\mathbf{A}_u\mathbf{A}_u^t = \mathbf{I}_{m-r}$. Then equality holds in (3) if and only if both (11) and (19) are equalities.

Consider equality in (19) which results from the application of Lemma 2 (inequality (15)) to $\mathbf{\Lambda} = T'(X^*)$. We have

$$\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t = \left(\begin{array}{c|c} \mathbf{\Lambda}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_u\mathbf{\Lambda}_u\mathbf{A}_u^t \end{array}\right) \tag{23}$$

where $\mathbf{\Lambda}_r = \mathrm{diag}(\lambda_1, \dots, \lambda_r)$ and $\mathbf{\Lambda}_u = \mathrm{diag}(\lambda_{r+1}, \dots, \lambda_n)$. Thus, we may choose $\mathbf{U}$ in the proof of Lemma 2 in the form $\mathbf{U} = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{U}_u \end{array}\right)$ where $\mathbf{U}_u$ is an $(m-r) \times (m-r)$ orthogonal matrix such that $\mathbf{U}_u\mathbf{A}_u\mathbf{\Lambda}_u\mathbf{A}_u^t\mathbf{U}_u^t$ is diagonal. Then $\mathbf{UA} = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{U}_u\mathbf{A}_u \end{array}\right)$ is still of the form (21) where $\mathbf{U}_u\mathbf{A}_u$ has orthonormal rows.

Therefore, equality in (15) is equivalent to equality in (17) where we may again assume that $\mathbf{A}$ is of the form (21) where r is maximal and $\mathbf{A}_u$ has orthonormal rows. By Remark 2, we may assume that all columns of $\mathbf{A}_u$ are nonzero. Notice that any row of $\mathbf{A}_u$ in (21) should have *at least two nonzero elements*. Otherwise, there would be one row of $\mathbf{A}_u$ of the form $(0, \dots, 0, \pm 1, 0, \dots, 0)$ with the nonzero element in the jth position. Since the rows are orthonormal, the other elements in the jth column would necessarily equal zero, and the corresponding component of X would be recoverable, which contradicts the maximality of r.

Now since the logarithm is strictly concave, equality holds in (17) if and only if for all $i = 1, 2, \dots, m$, all the λ_j for which $A_{i,j} \neq 0$ are equal. Because no column of $\mathbf{A}_u$ is zero and any row of $\mathbf{A}_u$ in (21) has at least two nonzero elements, this implies that for any j such that $r < j \le n$, λ_j is equal to another λ_k where $r < k \le n$, $k \neq j$. Since Lemma 2 was applied to $\mathbf{\Lambda} = T'(X^*)$ it follows that

$$T_j'(X_j^*) = T_k'(X_k^*) \text{ a.e.} \quad (r < j, k \le n) \tag{24}$$

Because X_j^* and X_k^* are independent, this implies that both $T_j'(X_j^*)$ and $T_k'(X_k^*)$ are constant and equal a.e., hence $T_j' = T_k' = c$ for some constant[1] c. Therefore T_j is linear and $X_j = T_j(X_j^*)$ is normal for all $r < j \le n$. This completes the proof.[2] □

[1] This is similar to what appeared in an earlier transportation proof of the EPI [5]. By (12), we necessarily have $c = 1$ if we assume that all individual entropies are equal as in § II-B.

[2] This implies, in particular, that equality in (19) implies equality in (11). This can also be seen directly: if $T_j' = 1$ for all $r < j \le n$, then for $\mathbf{A}$ of the form (21) in (11), $\mathbf{A}_u T(\mathbf{A}_u^t\widetilde{X} + \mathbf{A'}^t\widetilde{X}') = \widetilde{X}$ is independent of $\widetilde{X}'$.

IV. Extension to Complex Matrix and Variables

A complex random variable $X \in \mathbb{C}$ can always be viewed as a two-dimensional real random vector $\widehat{X} = \binom{\mathrm{Re}\, X}{\mathrm{Im}\, X} \in \mathbb{R}^2$. Therefore, by the vector form of the EPI [2]–[4], (1) holds for scalar coefficients $a_1, a_2 \in \mathbb{R}$ when $X_1, X_2 \in \mathbb{C}$ are independent complex random vectors and $X_1^*, X_2^* \in \mathbb{C}$ are independent white normal random vectors satisfying (2). Here "white normal" $X^* \in \mathbb{C}$ amounts to say that X^* is *proper* normal or *circularly symmetric* normal [13] (*c-normal* in short): $X^* \sim \mathcal{CN}(0, \sigma^2)$, that is, $\widehat{X^*} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_2)$.

That (1) also holds for *complex* coefficients $a_1, a_2 \in \mathbb{C}$ is less known but straightforward. To see this, define[3] $\widehat{a} = \binom{\mathrm{Re}\, a \;\; -\mathrm{Im}\, a}{\mathrm{Im}\, a \;\; \mathrm{Re}\, b}$ for any $a \in \mathbb{C}$, so that $\widehat{aX} = \widehat{a}\widehat{X}$. Then $h(aX) = h(\widehat{aX}) = h(\widehat{X}) + \log|\widehat{a}| = h(X) + \log|a|^2$. Hence (2) implies $h(a_1 X_1) = h(a_1 X_1^*)$ and $h(a_2 X_2) = h(a_2 X_2^*)$. In addition, if $X^* \sim \mathcal{CN}(0, \sigma^2)$ then $aX^* \sim \mathcal{CN}(0, |a|^2\sigma^2)$. Therefore, by the vector EPI applied to $a_1 X_1$ and $a_2 X_1$ we see that (1) holds for complex coefficients $a_1, a_2 \in \mathbb{C}$ when X_1^*, X_2^* are independent c-normal variables satisfying (2).

The extension of the matrix EPI (3) to *complex* $\mathbf{A}$ and X is more involved. We need the following notions (see, e.g., [14] and [15, chap. 10]). Define $\widehat{X} \in \mathbb{R}^{2n}$ by stacking the $\widehat{X_i}$ for each component $X_i \in \mathbb{C}$ of $X \in \mathbb{C}^n$, and define $\widehat{\mathbf{A}}$ as the $2m \times 2n$ real matrix with 2×2 entries $\widehat{A_{i,j}}$ where $A_{i,j}$ are the complex entries of $\mathbf{A}$. It is easily checked that $\widehat{\mathbf{A}X} = \widehat{\mathbf{A}}\widehat{X}$, $\widehat{\mathbf{AB}} = \widehat{\mathbf{A}}\widehat{\mathbf{B}}$, $\widehat{\mathbf{A}^\dagger} = \widehat{\mathbf{A}}^t$ where $\mathbf{A}^\dagger$ is the conjugate transpose, and $|\widehat{\mathbf{A}}| = |\mathbf{A}|^2$ where $|\mathbf{A}|$ denotes the modulus of the determinant of $\mathbf{A}$.

We also need the following extension of Lemma 1:

Lemma 4 (2D Brenier Map [16], [17]): *Let $\widehat{X^*} \in \mathbb{R}^2$ be a (white) normal random vector. For any given continuous density f over $\mathbb{R}^2$, there exists a differentiable transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ with symmetric positive definite Jacobian T' (noted $T' > 0$) such that $\widehat{X} = T(\widehat{X^*})$ has density f.*

Courtade et al. [18] noted that the Brenier map can be used in the transportation proof of [5] to prove Shannon's vector EPI. We find it also convenient to prove the complex matrix EPI:

Theorem 2: *The matrix EPI* (3) *holds for any $m \times n$ complex matrix $\mathbf{A}$ and any random vector X of independent complex components X_i, where X^* is a c-normal vector with independent components X_i^* satisfying* (4)*. If equality holds in* (3) *then all unrecoverable components present in $\mathbf{A}X$ (in the sense of Definition 1) are normal.*

The exact necessary and sufficient condition for equality is more involved and will appear elsewhere.

Proof: We sketch the proof by going through the above proofs in Sections II and III and pointing out the differences:

§II-A: The scaling property of entropy now reads $h(\mathbf{A}X) = h(\widehat{\mathbf{A}}\widehat{X}) = h(\widehat{X}) + \log|\widehat{\mathbf{A}}| = h(X) + \log|\mathbf{A}|^2$.

§II-B: Since $h(X^*) = \log\big(\pi e \sigma^2\big)$ for $X^* \sim \mathcal{CN}(0, \sigma^2)$, independent X_j^* with equal entropies are i.i.d.

§II-C: The Gram-Schmidt orthonormalization takes place in $\mathbb{C}^n$ with $h(\mathbf{A}'X) = h(\mathbf{A}X) + \log|\mathbf{L}|^2$.

[3]There is an ambiguity of notation easily resolved from the context: $\widehat{a}$ is a matrix when a is a constant and $\widehat{X}$ is a vector when X is random.

§II-D: $\mathbf{U} = \binom{\mathbf{A}}{\mathbf{A}'}$ is now an $n \times n$ *unitary* matrix. Recall that a circularly symmetric $X^* \sim \mathcal{CN}(0, \mathbf{K})$ is such that $\mathbf{A}X^* \sim \mathcal{CN}(0, \mathbf{AKA}^\dagger)$ for any $\mathbf{A}$. Since $X^* \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ is i.i.d., $\mathbf{U}X^* \sim \mathcal{CN}(0, \sigma^2\mathbf{UU}^\dagger = \sigma^2\mathbf{I})$ is also i.i.d. and the inverse transformation is the conjugate transpose $X^* = \mathbf{A}^\dagger\widetilde{X} + \mathbf{A}'^\dagger\widetilde{X}'$.

§II-E: Lemma 4 replaces Lemma 1 and (8) becomes

$$T'(\widehat{X^*}) = \mathrm{diag}\big(T_1'(\widehat{X_1^*}), \ldots, T_n'(\widehat{X_n^*})\big) \tag{25}$$

in *block-diagonal* form where each 2×2 block $T_i'(\widehat{X_i^*}) > 0$ is symmetric positive definite.

§II-G: In terms of the hat variables:

$$\mathbb{E}\log|T_j'(\widehat{X_j^*})| = 0 \qquad (j = 1, 2, \ldots, n). \tag{26}$$

where $|\cdot|$ denotes the absolute value of the determinant, and

$$\begin{aligned} &h\big(\widehat{\mathbf{A}}T(\widehat{\mathbf{A}}^t\widehat{\widetilde{X}} + \widehat{\mathbf{A}'}^t\widehat{\widetilde{X}'}) \mid \widehat{\widetilde{X}'}\big) \\ &\qquad = h(\widetilde{X}) + \mathbb{E}\log|\widehat{\mathbf{A}}T'(\widehat{X^*})\widehat{\mathbf{A}}^t| \end{aligned} \tag{27}$$

§II-H: We show that Lemma 2 still holds when $\mathbf{\Lambda}$ is block-diagonal with 2×2 diagonal blocks $\lambda_j > 0$ (symmetric positive definite). Write

$$\lambda_j = \widehat{u_j}\, d_j\, \widehat{u_j}^t \tag{28}$$

where d_j is 2×2 diagonal with positive diagonal elements and $\widehat{u_j}$ is a rotation matrix, corresponding to a complex unit $u_j = e^{i\theta_j}$. Then the block-diagonal $\widehat{\mathbf{U}} = \mathrm{diag}(\widehat{u_1}, \ldots, \widehat{u_n})$ is orthonormal and $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ is diagonal. We can now apply Lemma 2 to $\widehat{\mathbf{A}}\widehat{\mathbf{U}}$ and $\mathbf{D}$:

$$\log|\widehat{\mathbf{A}}\mathbf{\Lambda}\widehat{\mathbf{A}}^t| \geq \mathrm{tr}(\widehat{\mathbf{A}}\widehat{\mathbf{U}}[\log\mathbf{D}]\widehat{\mathbf{U}}^t\widehat{\mathbf{A}}^t) \tag{29}$$

$$= \mathrm{tr}(\widehat{\mathbf{A}}[\log\mathbf{\Lambda}]\widehat{\mathbf{A}}^t) \tag{30}$$

where $\log\mathbf{\Lambda}$ is the (block diagonal) logarithm of $\mathbf{\Lambda} > 0$. Thus

$$\mathrm{tr}(\widehat{\mathbf{A}}[\log\mathbf{\Lambda}]\widehat{\mathbf{A}}^t) = \sum\nolimits_i \mathrm{tr}\big(\sum\nolimits_j \widehat{A_{i,j}}[\log\lambda_j]\widehat{A_{i,j}}^t\big) \tag{31}$$

$$= \sum\nolimits_i \sum\nolimits_j |A_{i,j}|^2 \mathrm{tr}(\log\lambda_j) \tag{32}$$

where $\mathrm{tr}(\log\lambda_j) = \log|\lambda_j|$ since λ_j is symmetric positive definite. Thus we obtain

$$\mathbb{E}\log|\widehat{\mathbf{A}}T'(X^*)\widehat{\mathbf{A}}^t| \geq \sum\nolimits_i \sum\nolimits_j |A_{i,j}|^2 \mathbb{E}\log|T_j'(\widehat{X_j^*})| = 0 \tag{33}$$

which is the final step to prove the (complex) matrix EPI (3).

Assume that equality holds in (3) as in the converse part of the proof of Theorem 1 (Section III). That proof is unchanged up to the point where one considers the equality condition in Lemma 2 applied to $\widehat{\mathbf{A}}\widehat{\mathbf{U}}$ and diagonal $\mathbf{D}$, that is, in (29). By the strict concavity of the logarithm, equality holds in (29) if and only if for any two nonzero elements in the same row of $\widehat{\mathbf{A}}\widehat{\mathbf{U}} = \widehat{\mathbf{AU}}$, the corresponding two diagonal elements of $\mathbf{D}$ are equal. Since $\mathbf{U} = \mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$, the nonzero elements of $\mathbf{AU}$ are at the same places as those of $\mathbf{A}$, where $\mathbf{A}$ is of the form (21). Therefore, due to the structure of $\widehat{\mathbf{AU}}$, for any j such that $r < j \leq n$, the two diagonal elements of d_j are equal to the two diagonal elements of another d_k where $r < k \leq n$, $k \neq j$, which implies $\lambda_j = \lambda_k$. This gives (24) from which one concludes as before that for all $r < j \leq n$, T_j is linear, and, therefore, $X_j = T_j(X_j^*)$ is normal. □

V. Application to Blind Source Extraction

The theoretical setting of the blind source extraction problem is as follows [9]. We are given n (zero-mean) independent (real or complex) "sources" $X = (X_1, X_2, \ldots, X_n)^t$ which are mixed using an $n \times n$ invertible (real or complex) matrix $\mathbf{M}$, resulting in the observation $Y = \mathbf{M}X$. The covariance matrix $\mathbf{K}_Y$ of Y can be estimated but both $\mathbf{M}$ and X are unknown. Since one can introduce arbitrary scaling factors in $\mathbf{M}$ and X for the same observation Y, we can assume an arbitrary normalization of the sources. For convenience we assume here that they have the same entropies:

$$h(X_1) = h(X_2) = \cdots = h(X_n). \tag{34}$$

Blind source extraction (or partial BSS) of m sources ($1 \leq m \leq n$) aims at finding a (full rank) $m \times n$ matrix $\mathbf{W}$ such that $Z = \mathbf{W}Y$ is composed of m (out of n) original sources, up to order and scaling. In other words $\mathbf{A} = \mathbf{W}\mathbf{M}$ should have exactly one nonzero element per row.

Definition 2 (Contrast function [9]): *A* contrast $\mathcal{C}(\mathbf{W})$ *is a function that is invariant to permutation and scaling of the rows* $\mathbf{w}_i$ *of* $\mathbf{W}$*, and such that it achieves a* minimum *if only if* $\mathbf{A} = \mathbf{W}\mathbf{M}$ *has one nonzero element per row.*

Theorem 3: *Assume that* at most one *source is normal. Then*

$$\mathcal{C}(\mathbf{W}) = \sum_{i=1}^{m} h(\mathbf{w}_i Y) - \frac{1}{2}\log|\mathbf{W}\mathbf{K}_Y\mathbf{W}^t| \tag{35}$$

where $\mathbf{w}_i$ *are the rows of* $\mathbf{W}$*, is a contrast function.*

Such a contrast function was first proposed by Pham [19] (see also [20]) in the real case with a different proof that uses the classical EPI for $m = 1$ and Hadamard's inequality. It is particularly interesting to rewrite it in terms of the matrix EPI:

Proof: The real and complex cases being similar, we prove the result in the real case. Let $\mathbf{A} = \mathbf{W}\mathbf{M}$ and let X^* be as in (3). For i.i.d. components we can rewrite [6, Eq. (13)] as $h(\mathbf{A}X^*) = mh + \frac{1}{2}\log|\mathbf{A}\mathbf{A}^t|$ where h is the common value of (34). Since $Z = \mathbf{W}Y = \mathbf{A}X$, up to an additive constant we may decompose $\mathcal{C}$ as

$$\mathcal{C}(\mathbf{W}) = \mathcal{C}_h(\mathbf{W}) + \mathcal{C}_i(\mathbf{W}) + \mathrm{Cst}. \tag{36}$$

where

$$\mathcal{C}_h(\mathbf{W}) = h(\mathbf{A}X) - h(\mathbf{A}X^*) \geq 0 \tag{37}$$

$$\mathcal{C}_i(\mathbf{W}) = \sum_i h(Z_i) \; - \; h(Z) \geq 0 \tag{38}$$

The term $\mathcal{C}_i(\mathbf{W})$ is minimum (with minimum value $= 0$) if and only if the components Z_i of Z are independent.

The $\mathcal{C}_h(\mathbf{W})$ is minimum (with minimum value $= 0$) if and only if equality holds in (3). Since at most one source is normal, at most one source present in $\mathbf{A}X$ can be unrecoverable. But if one (normal) source is not recoverable, the canonical form (21) implies that at most one column of $\mathbf{A}_u$ is nonzero, which contradicts the maximality of r in Lemma 3. Therefore, $r = m$ and the canonical form of $\mathbf{A}$ becomes $(\mathbf{I}_m \mid 0)$.

With the additional constraint $\mathcal{C}_i(\mathbf{W}) = 0$ that components of $Z = \mathbf{A}X$ are independent, it follows from the Darmois–Skitovich theorem [21] (see [14] in the complex case) that $\mathbf{A}$ has exactly one nonzero per row. □

Interestingly, the contrast function in the form (36) represents a transition between the two well-known extreme cases:

- $m = 1$, for which $\mathcal{C}_i = 0$ where each source is extracted one by one using the classical EPI (minimize $\mathcal{C}_h$);
- $m = n$, for which $\mathcal{C}_h = 0$, where all n sources are separated simultaneously; we are then reduced to an independent component analysis (ICA) problem [14], [21] in which the multivariate "mutual information" $\mathcal{C}_i = D(p(Z)\|\prod_i p(Z_i))$ is minimized.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, Oct. 1948.

[2] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Commun. Math. Phys.*, vol. 62, pp. 35–41, 1978.

[3] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[4] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.

[5] ——, "Yet another proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3595–3599, Jun. 2017.

[6] R. Zamir and M. Feder, "A generalization of the entropy power inequality with applications," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1723–1728, Sep. 1993.

[7] ——, "A generalization of information theoretic inequalities to linear transformations of independent vector," in *Proc. Sixth Joint Swedish-Russian International Workshop on Information Theory*, Mölle, Sweden, Aug. 1993, pp. 254–258.

[8] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, USA, Jul. 2006, pp. 1011–1015.

[9] F. Vrins, *Contrast properties of entropic criteria for blind source separation: A unifying framework based on information-theoretic inequalities*. Louvain University Press (UCL), Mar. 2007.

[10] J.-F. Cardoso, "An efficient technique for the blind separation of complex sources," in *Proc. IEEE Signal Proc. Workshop on Higher-Order Statistics*, South Lake Tahoe, CA, Jun. 1993, pp. 275–279.

[11] O. Rioul, "Optimal transportation to the entropy-power inequality," in *IEEE Information Theory and Applications Workshop (ITA 2017)*, San Diego, USA, Feb. 2017.

[12] R. Zamir and M. Feder, "On the volume of the Minkowski sum of line sets and the entropy-power inequality," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 3039–3043, Nov. 1998.

[13] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[14] 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.

[15] O. Rioul, *Théorie des probabilités [in French]*. London, UK: Hermes Science - Lavoisier, 2008.

[16] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Commun. Pure Appl. Math.*, vol. 44, no. 4, pp. 375–417, Jun. 1991.

[17] R. J. McCann, "Existence and uniqueness of monotone measure-preserving maps," *Duke Math. J.*, vol. 80, no. 2, pp. 309–324, 1995.

[18] T. A. Courtade, M. Fathi, and A. Pananjady, "Quantitative stability of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5691–5703, Aug. 2018.

[19] D.-T. Pham, "Blind partial separation of instantaneous mixtures of sources," in *Proc. 6th International Conference on Independent Component Analysis (ICA)*. Charleston, SC, USA: Springer, March 5–8 2006, pp. 37–42.

[20] S. Cruces, A. Cichocki, and S. Amari, "The minimum entropy and cumulants based contrast functions for blind source extraction," in *6th International Work-Conference on Artificial Neural Networks (IWANN)*. Granada, Spain: Springer, 2001, pp. 786–793.

[21] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.