Comments and Corrections

Counterexample to the Vector Generalization of Costa's Entropy Power Inequality, and Partial Resolution

Thomas A. Courtade, Guangyue Han, and Yaochen Wu

***Abstract*— We give a counterexample to the vector generalization of Costa's entropy power inequality due to Liu *et al.* In particular, the claimed inequality can fail if the matrix-valued parameter in the convex combination does not commute with the covariance of the additive Gaussian noise. Conversely, the inequality holds if these two matrices commute.**

***Index Terms*— Entropy power inequality, Costa's EPI.**

I. Introduction and Main Result

For a random vector X with density on $\mathbb{R}^n$, let $h(X)$ denote its differential entropy. Let $Z \sim N(0, \Sigma_Z)$ be a Gaussian vector in $\mathbb{R}^n$ independent of X, and let A be a (real symmetric) positive semidefinite $n \times n$ matrix satisfying $\mathrm{A} \preceq \mathrm{I}$ with respect to the positive semidefinite ordering, where I denotes the identity matrix. In [1, Th. 1], Liu *et al.* claim the following generalization of Costa's EPI[1] [2]:

$$e^{\frac{2}{n}h\ (X+\mathrm{A}^{1/2}Z)} \geq |\mathrm{I}-\mathrm{A}|^{1/n} e^{\frac{2}{n}h(X)} + |\mathrm{A}|^{1/n} e^{\frac{2}{n}h(X+Z)}. \qquad (1)$$

Liu *et al.* apply (1) in their investigation of the secrecy capacity region of the degraded vector Gaussian broadcast channel with layered confidential messages.

The purpose of this note is to demonstrate that (1) can fail for $n \geq 2$, and also to offer a partial resolution. Toward the first goal, consider $n = 2$ and let us define

$$\Sigma_X = \begin{pmatrix} 200 & 100 \\ 100 & 51 \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} 200 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\mathrm{A}^{1/2} = \frac{1}{20}\begin{pmatrix} 10 & 5 \\ 5 & 17 \end{pmatrix}. \qquad (2)$$

Taking $X \sim N(0, \Sigma_X)$ and $Z \sim N(0, \Sigma_Z)$ to be independent Gaussian vectors, we have

$$\tfrac{1}{2\pi e} e^{\frac{2}{n}h\ (X+\mathrm{A}^{1/2}Z)} = |\Sigma_X + \mathrm{A}^{1/2}\Sigma_Z \mathrm{A}^{1/2}|^{1/2} \approx 19.53.$$

On the other hand,

$$\tfrac{1}{2\pi e}\left(|\mathrm{I}-\mathrm{A}|^{1/n} e^{\frac{2}{n}h(X)} + |\mathrm{A}|^{1/n} e^{\frac{2}{n}h(X+Z)}\right)$$
$$= |\mathrm{I}-\mathrm{A}|^{1/2}|\Sigma_X|^{1/2} + |\mathrm{A}|^{1/2}|\Sigma_X + \Sigma_Z|^{1/2} \approx 40.28.$$

Thus, a contradiction to (1) is obtained. We remark that there is nothing particularly unique about this counterexample, except that the matrices were chosen to violate (1) by a significant margin.

Evidently, further assumptions are needed in order for (1) to hold. To give a simple resolution, we note that it suffices for the matrices A and Σ_Z to commute.

Theorem 1: Let X be a random vector with density on $\mathbb{R}^n$ whose entropy exists in the usual Lebesgue sense, and let $Z \sim N(0, \Sigma_Z)$ be a Gaussian vector in $\mathbb{R}^n$ independent of X. If $\mathrm{A} \preceq \mathrm{I}$ is positive semidefinite and commutes with Σ_Z, then

$$e^{\frac{2}{n}h\ (X+\mathrm{A}^{1/2}Z)} \geq |\mathrm{I}-\mathrm{A}|^{1/n} e^{\frac{2}{n}h(X)} + |\mathrm{A}|^{1/n} e^{\frac{2}{n}h(X+Z)}.$$

Proof: For brevity, we refer the reader to the original proof of [1, Th. 1], and only point out where the argument needs to be corrected. To this end, Liu *et al.*'s proof contains an incorrect application of the AM-GM inequality in the form [1, eq. (28)]:

$$|\Sigma_Z^{-1}\,\mathsf{Cov}(Z|\mathrm{D}_\gamma X + Z)(\mathrm{I}-\mathrm{D}_\gamma^{-2})|^{1/n}$$
$$\leq \frac{1}{n}\mathsf{Tr}(\Sigma_Z^{-1}\,\mathsf{Cov}(Z|\mathrm{D}_\gamma X + Z)(\mathrm{I}-\mathrm{D}_\gamma^{-2})), \qquad (3)$$

where $\mathrm{D}_\gamma := (\mathrm{I}+\gamma(\mathrm{A}-\mathrm{I}))^{1/2}$, and $\gamma \in [0, 1]$ parameterizes a path of perturbation. Indeed, a product of positive semidefinite matrices is not necessarily positive semidefinite, which can lead to failure of the AM-GM inequality in the form (3). For example, returning to the counterexample above where $X \sim N(0, \Sigma_X)$ and matrices are chosen according to (2), the eigenvalues of $\Sigma_Z^{-1}\,\mathsf{Cov}(Z|\mathrm{D}_\gamma X + Z)(\mathrm{I}-\mathrm{D}_\gamma^{-2})$ can be approximately computed as $\{-0.0053, -0.7273\}$ for $\gamma = 0.5$, in violation of (3).

However, if A and Σ_Z commute, then so do $\Sigma_Z^{-1/2}$ and $(\mathrm{I}-\mathrm{D}_\gamma^{-2})^{1/2}$ since real symmetric matrices commute if and only if they are simultaneously diagonalizable by some orthogonal matrix U. Hence,

$$\mathsf{Tr}(\Sigma_Z^{-1}\,\mathsf{Cov}(Z|\mathrm{D}_\gamma X + Z)(\mathrm{I}-\mathrm{D}_\gamma^{-2}))$$
$$= \mathsf{Tr}((\mathrm{I}-\mathrm{D}_\gamma^{-2})^{1/2}\Sigma_Z^{-1/2}\mathsf{Cov}(Z|\mathrm{D}_\gamma X + Z)\Sigma_Z^{-1/2}(\mathrm{I}-\mathrm{D}_\gamma^{-2})^{1/2}).$$

The argument of the second trace term is clearly positive semidefinite, and therefore (3) holds for all $\gamma \in [0, 1]$ under the additional assumption that A and Σ_Z commute, thereby repairing Liu *et al.*'s proof. □

Manuscript received June 13, 2017; revised February 13, 2018; accepted April 6, 2018. Date of publication May 4, 2018; date of current version June 20, 2018. T. A. Courtade was supported by NSF under Grant CCF-1704967. G. Han was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project 17301017.

T. A. Courtade is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: courtade@berkeley.edu).

G. Han and Y. Wu are with the Department of Mathematics, The University of Hong Kong, Hong Kong.

Communicated by M. Costa, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2018.2833421

[1]Entropies are taken to be base e throughout. For a positive semidefinite matrix M, we write $\mathrm{M}^{1/2}$ to denote the unique positive semidefinite matrix such $\mathrm{M} = \mathrm{M}^{1/2}\mathrm{M}^{1/2}$.

II. Remarks

The critical application of inequality (1) by Liu *et al.* (see [1, p. 1877]) assumes only that A and Σ_Z are diagonal matrices, so the conclusions of [1] appear to be unaffected, aside from [1, Th. 1]. In fact, Ekrem and Ulukus [3] obtained the same secrecy capacity results using a different argument. Nevertheless, [1] has been cited numerous times in the literature, so other published results may be affected to varying degrees. As one example, a computation in [4] similarly overlooks non-commutativity of the matrices A and Σ_Z, leading to the incorrect conclusion that [1, Th. 1] is a corollary of [4, Th. 3]. Despite the error, the validity of [4, Th. 3] is not impacted, and the corrected computation, showing that [4, Th. 3] implies Theorem 1 above, can be found in [5, Sec. III.A]. Other works may be more seriously affected, but we do not attempt to give an accounting of consequences here.

In closing, we remark that the additional assumption that A and Σ_Z commute is a relatively strong one. It can easily be seen, using the simultaneous diagonalization property of A and Σ_Z by a common orthogonal matrix, that Theorem 1 has a completely equivalent statement where $Z \sim N(0, \mathrm{I})$ and A is restricted to be a diagonal matrix with diagonal entries $0 \le a_i \le 1$, $i = 1, \dots, n$. Theorem 1 should be viewed as an extension of Costa's original 1985 result (which assumed identical parameters a_i) in this sense. As pointed out to the authors by an anonymous referee, a standard information-theoretic argument may be used to establish Theorem 1 as a corollary of Costa's original inequality, without appealing to the perturbation framework used in [1]. This argument has been included in the appendix.

Appendix

What follows is a proof of Theorem 1 using Costa's entropy power inequality [2]. In particular, we shall prove Theorem 1 in the equivalent setting noted in the closing statement above, where $Z \sim N(0, \mathrm{I})$ and A is a diagonal matrix with diagonal entries $0 \le a_i \le 1$, $i = 1, \dots, n$. The argument below was provided by an anonymous referee.

$$Y_1 := X + \mathrm{A}^{1/2} Z_1, \text{ and } Y_2 := Y_1 + (\mathrm{I} - \mathrm{A})^{1/2} Z_2,$$

where Z_1, Z_2 are independent copies of $Z \sim N(0, \mathrm{I})$, also independent of X. Since A is assumed diagonal with entries $a_1, \dots, a_n$, Theorem 1 may thus be written as

$$\prod_{i=1}^{n} (1 - a_i)^{1/n} e^{\frac{2}{n}(h(X) - h(Y_1))} + \prod_{i=1}^{n} a_i^{1/n} e^{\frac{2}{n}(h(Y_2) - h(Y_1))} \le 1. \tag{4}$$

We consider the exponential terms separately. By the Csiszár sum identity,

$$\begin{aligned} h(X) - h(Y_1) &= \sum_{i=1}^{n} \Big(h(X_i | X^{i-1}, Y_{1,i+1}^n) \\ &\quad - h(Y_{1,i} | X^{i-1}, Y_{1,i+1}^n) \Big) \\ &= \sum_{i=1}^{n} \left(h(X_i | V_i) - h(Y_{1,i} | V_i) \right), \end{aligned}$$

where $V_i := (X^{i-1}, Y_{1,i+1}^n)$. Similarly,

$$\begin{aligned} h(Y_2) - h(Y_1) &= \sum_{i=1}^{n} \Big(h(Y_{2,i} | Y_2^{i-1}, Y_{1,i+1}^n) \\ &\quad - h(Y_{1,i} | Y_2^{i-1}, Y_{1,i+1}^n) \Big) \\ &\le \sum_{i=1}^{n} \left(h(Y_{2,i} | V_i) - h(Y_{1,i} | V_i) \right), \end{aligned}$$

where the last inequality follows from

$$I(X^{i-1}; Y_{2,i} | Y_2^{i-1}, Y_{1,i+1}^n) \le I(X^{i-1}; Y_{1,i} | Y_2^{i-1}, Y_{1,i+1}^n),$$

which holds by Markovity induced by construction of Y_1, Y_2. Hence, (4) holds if

$$e^{\frac{1}{n}\sum_{i=1}^{n} \log(1-a_i) + 2\ h(X_i|V_i) - 2\ h(Y_{1,i}|V_i)} + e^{\frac{1}{n}\sum_{i=1}^{n} \log(a_i) + 2\ h(Y_{2,i}|V_i) - 2\ h(Y_{1,i}|V_i)} \le 1.$$

By convexity of $x \mapsto e^x$, it is sufficient to show that

$$(1 - a_i) e^{2\ h(X_i|V_i) - 2\ h(Y_{1,i}|V_i)} + a_i e^{2\ h(Y_{2,i}|V_i) - 2\ h(Y_{1,i}|V_i)} \le 1$$

for each i. However, this is just the conditional form of Costa's scalar inequality for concavity of entropy power, which holds again by convexity of $x \mapsto e^x$, and the easily verified fact that $V_i \to X_i \to Y_{1,i} \to Y_{2,i}$.

References

[1] R. Liu, T. Liu, H. V. Poor, and S. Shamai (Shitz), "A vector generalization of Costa's entropy-power inequality with applications," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1865–1879, Apr. 2010.

[2] M. Costa, "A new entropy power inequality," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 751–760, Nov. 1985.

[3] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2083–2114, Apr. 2011.

[4] T. A. Courtade, "Strengthening the entropy power inequality," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 2294–2298.

[5] T. A. Courtade, "A strong entropy power inequality," in *Proc. IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2173–2192, Apr. 2018.

Thomas A. Courtade received the B.Sc. degree *(summa cum laude)* in electrical engineering from Michigan Technological University in 2007, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA) in 2008 and 2012, respectively. He is currently an Assistant Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley.

Prof. Courtade's honors include an NSF CAREER award, a Hellman Fellowship, the Distinguished Ph.D. Dissertation Award and an Excellence in Teaching Award from the UCLA Department of Electrical Engineering, and a Jack Keil Wolf Student Paper Award for the 2012 International Symposium on Information Theory.

Guangyue Han received the B.S. and M.S. degrees in mathematics from Peking University, China, and the Ph.D. degree in mathematics from the University of Notre Dame, U.S.A. in 1997, 2000 and 2004, respectively. After three years with the Department of Mathematics at the University of British Columbia, Canada, he joined the Department of Mathematics at the University of Hong Kong, China in 2007. His main research areas are coding and information theory.

Yaochen Wu is currently an undergraduate student majoring in mathematics at the University of Hong Kong. He expects to receive the B.S. degree in mathematics in August 2018.