Correspondence

A Simple Proof of the Entropy-Power Inequality

Sergio Verdú, *Fellow, IEEE,* and Dongning Guo, *Member, IEEE*

***Abstract*—This correspondence gives a simple proof of Shannon's entropy-power inequality (EPI) using the relationship between mutual information and minimum mean-square error (MMSE) in Gaussian channels.**

***Index Terms*—Differential entropy, entropy-power inequality (EPI), minimum mean-square error (MMSE).**

I. Introduction

In *A Mathematical Theory of Communication* [1], Claude Shannon put forth the inequality

$$\exp(2h(X+Y)) \geq \exp(2h(X)) + \exp(2h(Y)) \tag{1}$$

where X and Y are independent real-valued random variables and $h(X)$ is the differential entropy of the probability density function f_X[1]

$$h(X) = -\int_{-\infty}^{\infty} f_X(u) \log f_X(u)\, du.$$

The entropy-power (variance of a Gaussian random variable with the same differential entropy) is maximum and equal to the variance when the random variable is Gaussian, and thus, the essence of (1) is that the sum of independent random variables tends to be "more Gaussian" than one or both of the individual components. Note that (1) is equivalent to

$$\exp(2h(X_1 + \cdots + X_n)) \geq \sum_{i=1}^{n} \exp(2h(X_i)) \tag{2}$$

for n independent random variables.

The first proof of (1) was given by Stam [2], based on an identity communicated to him by N. G. De Bruijn, which couples Fisher's information with Shannon's differential entropy.

Capitalizing on the relationship between mutual information and minimum mean-square error (MMSE) for additive Gaussian channels [3], this note gives a simpler proof of the entropy-power inequality (EPI) based on an elementary estimation-theoretic reasoning which sidesteps invoking Fisher's information. In a follow-up to this work [4], we use the MMSE to give simple proofs of two variations of the EPI, namely, Costa's strengthened EPI in which one of the variables is Gaussian [5], and a generalized EPI for linear transforms of a random vector due to Zamir and Feder [6].

Following a simple "noise-incremental" argument, [3] shows that regardless of the distribution of X we can write

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} I(X; \sqrt{\gamma}X + N) = \frac{1}{2}\mathsf{mmse}(X, \gamma) \tag{3}$$

where $N \sim \mathcal{N}(0,1)$ is standard Gaussian independent of X, and the MMSE of estimating X in unit-variance additive Gaussian noise is

$$\mathsf{mmse}(X, \gamma) = \mathsf{E}\left\{(X - \mathsf{E}\{X \mid \sqrt{\gamma}X + N\})^2\right\}.$$

Here γ is understood as the (gain of the) signal-to-noise ratio of the Gaussian channel whose input is X.

A direct consequence of (3) is the representation of the differential entropy of a random variable with variance σ_X^2 as [3, eq. (182)]

$$h(X) = \frac{1}{2}\log\left(2\pi e\, \sigma_X^2\right) - \frac{1}{2}\int_0^{\infty} \frac{\sigma_X^2}{1+\gamma\,\sigma_X^2} - \mathsf{mmse}(X,\gamma)\, \mathrm{d}\gamma. \tag{4}$$

Thus, the nongaussianness of X (divergence of f_X with respect to the Gaussian density with identical first and second moments) is given by one half of the integral of the difference of the MMSEs achievable by a Gaussian input with variance σ_X^2 and by X, respectively.

For a unit-variance X, (4) reduces to

$$h(X) = \frac{1}{2}\log(2\pi e) - \frac{1}{2}\int_0^{\infty} \frac{1}{1+\gamma} - \mathsf{mmse}(X,\gamma)\, \mathrm{d}\gamma. \tag{5}$$

It is amusing (and useful) to note that (5) holds even if X does not have unit variance: simply observe that

$$\log \sigma_X^2 = \int_0^{\infty} \frac{\sigma_X^2}{1+\gamma\,\sigma_X^2} - \frac{1}{1+\gamma}\, \mathrm{d}\gamma. \tag{6}$$

Note that whenever $\mathsf{mmse}(X,\gamma) = o(1/\gamma)$ (as in the case of a discrete random variable, where it vanishes exponentially), (4) indicates that $h(X) = -\infty$.

Since (5) expresses the differential entropy of an arbitrary random variable in terms of the MMSE of its estimation when observed in Gaussian noise, (1) can be seen as a relationship between the MMSEs (integrated over signal-to-noise ratios) of the sum and of the individual random variables.

II. Proof of (1)

Instead of showing (1) directly, it is more convenient to prove the equivalent inequality:

Lemma 1: (Lieb [7])[2] The inequality

$$h(X_1 \cos\alpha + X_2 \sin\alpha) \geq \cos^2\alpha\, h(X_1) + \sin^2\alpha\, h(X_2) \tag{7}$$

for all independent X_1 and X_2 and $\alpha \in [0, 2\pi]$ is equivalent to the EPI (1) for all independent X and Y.

Manuscript received July 13, 2005. This work was supported in part by the National Science Foundation under Grants NCR-0074277 and CCR-0312879.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: Verdu@Princeton.EDU).

D. Guo is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA (e-mail: dGuo@Northwestern.EDU).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2006.872978

[1]For convenience, throughout this correspondence we assume that all logarithms are natural.

[2]See the Appendix for the proof.

Proof: [Inequality (7)]. Select arbitrary independent X_1 and X_2, $\alpha \in [0, 2\pi]$. According to (5), the difference between the left- and right-hand sides of (7) is equal to

$$\frac{1}{2}\int_0^\infty \mathsf{mmse}\left(X_1\cos\alpha + X_2\sin\alpha, \gamma\right) - \cos^2\alpha\,\mathsf{mmse}\left(X_1,\gamma\right) - \sin^2\alpha\,\mathsf{mmse}\left(X_2,\gamma\right)\mathrm{d}\gamma.$$

Thus, it suffices to show that for all γ

$$\mathsf{mmse}\left(X,\gamma\right) \geq \cos^2\alpha\,\mathsf{mmse}\left(X_1,\gamma\right) + \sin^2\alpha\,\mathsf{mmse}\left(X_2,\gamma\right) \quad (8)$$

where we define

$$X = X_1\cos\alpha + X_2\sin\alpha.$$

Let N_1 and N_2 be independent standard Gaussian random variables, and let

$$\begin{aligned} Z_1 &= \sqrt{\gamma}\,X_1 + N_1 \\ Z_2 &= \sqrt{\gamma}\,X_2 + N_2 \\ Z &= Z_1\cos\alpha + Z_2\sin\alpha. \end{aligned}$$

Then, the left-hand side of (8) can be written as

$$\begin{aligned} &\mathsf{E}\left\{\left(X - \mathsf{E}\left\{X \mid Z\right\}\right)^2\right\} \\ &\quad\geq \mathsf{E}\left\{\left(X - \mathsf{E}\left\{X \mid Z_1, Z_2\right\}\right)^2\right\} \quad (9) \\ &\quad= \cos^2\alpha\,\mathsf{E}\left\{\left(X_1 - \mathsf{E}\left\{X_1 \mid Z_1\right\}\right)^2\right\} \\ &\qquad+ \sin^2\alpha\,\mathsf{E}\left\{\left(X_2 - \mathsf{E}\left\{X_2 \mid Z_2\right\}\right)^2\right\} \quad (10) \end{aligned}$$

thereby showing (8). Note that in (10) we have used the mutual independence of N_1, N_2, X_1, X_2. □

Note that the only inequality used in the proof of (7) (and, hence, (1) via Lemma 1) is (9), namely, the fact that it is easier to estimate the sum of independent random variables on the basis of individual measurements than on the basis of their sum.

As an addendum to the foregoing proof, we call attention to the fact that the main inequality (7) also holds when the random variables therein are finitely or countably (real) valued and the differential entropies are replaced by entropies. To see this, note that in the discrete case, the representation

$$\begin{aligned} &H(X_1\cos\alpha + X_2\sin\alpha) - \cos^2\alpha\,H(X_1) - \sin^2\alpha\,H(X_2) \\ &= \int_0^\infty \mathsf{mmse}\left(X_1\cos\alpha + X_2\sin\alpha,\gamma\right) \\ &\quad - \cos^2\alpha\,\mathsf{mmse}\left(X_1,\gamma\right) - \sin^2\alpha\,\mathsf{mmse}\left(X_2,\gamma\right)\mathrm{d}\gamma \end{aligned}$$

follows immediately from [3, eq. (176)], and the argument we gave above for (8) holds also for discrete random variables.

Unlike previous proofs of the EPI (e.g., [2], [8]), the new proof does not hinge on Fisher's information.

III. Vector Entropy-Power Inequality

To show the vector EPI for independent random n-vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$,

$$\exp\left(2h(\boldsymbol{X}+\boldsymbol{Y})/n\right) \geq \exp\left(2h(\boldsymbol{X})/n\right) + \exp\left(2h(\boldsymbol{Y})/n\right)$$

we can follow the same steps as above using the following representation for the differential entropy of an n vector with covariance matrix $\boldsymbol{\Sigma}$, which follows easily from [3, eq. (22)]:

$$\begin{aligned} h(\boldsymbol{X}) &= \frac{1}{2}\log\left((2\pi e)^n\,\det\boldsymbol{\Sigma}\right) \\ &\quad - \frac{1}{2}\int_0^\infty \mathsf{tr}\left[\boldsymbol{\Sigma}^{-1} + \gamma\mathbf{I}\right]^{-1} - \mathsf{mmse}\left(\boldsymbol{X},\gamma\right)\mathrm{d}\gamma \\ &= \frac{n}{2}\log\left(2\pi e\right) - \frac{1}{2}\int_0^\infty \frac{n}{1+\gamma} - \mathsf{mmse}\left(\boldsymbol{X},\gamma\right)\mathrm{d}\gamma. \quad (11) \end{aligned}$$

Here, the MMSE of estimating a vector in Gaussian noise is

$$\mathsf{mmse}\left(\boldsymbol{X},\gamma\right) = \mathsf{E}\left\{\left\|\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X} \mid \sqrt{\gamma}\,\boldsymbol{X} + \boldsymbol{N}\right\}\right\|^2\right\}$$

where $\boldsymbol{N}$ consists of independent and identically distributed (i.i.d.) unit-variance Gaussian entries.

With (11), the same proof we used above can be employed to show that (7) holds when X_1 and X_2 therein are replaced by independent vectors of identical dimensions.

Appendix

For completeness we include a simple proof of Lemma 1.

Proof [Lemma 1]

To verify that (1) follows from (7), choose arbitrary independent X and Y and let

$$\tan\alpha = \exp\left(h(Y) - h(X)\right) \quad (12)$$

$$X_1 = \frac{X}{\cos\alpha} \quad (13)$$

$$X_2 = \frac{Y}{\sin\alpha}. \quad (14)$$

Then, using (7) and $h(aV) = h(V) + \log|a|$, we can bound

$$\begin{aligned} &h(X+Y) \\ &\geq \cos^2\alpha\left(h(X) - \log\cos\alpha\right) + \sin^2\alpha\left(h(Y) - \log\sin\alpha\right) \\ &= \frac{1}{2}\log\left[\exp(2h(X)) + \exp(2h(Y))\right] \end{aligned}$$

which is (1).

The reverse direction is not needed in the proof of (1), but it is very simple: fixing arbitrary α and independent X_1, X_2, choose X and Y to satisfy (13) and (14). Upon taking logarithms of both sides of (1), (7) follows from the concavity of the logarithm. □

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul./Oct. 1948.

[2] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Contr.*, vol. 2, pp. 101–112, 1959.

[3] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[4] ——, "Proof of entropy power inequalities via MMSE," in *Proc. 2006 IEEE Int. Symp. Information Theory*, 2006, submitted for publication.

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

[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] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Commun. Math. Phys.*, vol. 62, no. 1, pp. 35–41, 1978.

[8] 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.