Correspondence

On the Similarity of the Entropy Power Inequality and the Brunn–Minkowski Inequality

MAX H.M. COSTA MEMBER, IEEE, AND
THOMAS M. COVER FELLOW, IEEE

***Abstract*—The entropy power inequality states that the effective variance (entropy power) of the sum of two independent random variables is greater than the sum of their effective variances. The Brunn–Minkowski inequality states that the effective radius of the set sum of two sets is greater than the sum of their effective radii. Both these inequalities are recast in a form that enhances their similarity. In spite of this similarity, there is as yet no common proof of the inequalities. Nevertheless, their intriguing similarity suggests that new results relating to entropies from known results in geometry and vice versa may be found. Two applications of this reasoning are presented. First, an isoperimetric inequality for entropy is proved that shows that the spherical normal distribution minimizes the trace of the Fisher information matrix given an entropy constraint—just as a sphere minimizes the surface area given a volume constraint. Second, a theorem involving the effective radii of growing convex sets is proved.**

I. The Entropy Power Inequality

Let a random variable X have a probability density function $f(x)$, $x \in \mathbf{R}$. Then its (differential) *entropy* $H(X)$ is defined as

$$H(X) = -\int f(x)\ln f(x)\,dx.$$

Shannon's entropy power inequality [1] states that for X and Y independent random variables having density functions

$$e^{2H(X+Y)} \geq e^{2H(X)} + e^{2H(Y)}. \tag{1}$$

We wish to recast this inequality. First we observe that a normal random variable $Z \sim \phi(z) = (1/\sqrt{2\pi\sigma^2})e^{-z^2/2\sigma^2}$ with variance σ^2 has entropy

$$H(Z) = -\int \phi \ln \phi = \frac{1}{2}\ln 2\pi e\sigma^2. \tag{2}$$

By inverting, we see that if Z is normal with entropy $H(Z)$, then its variance is

$$\sigma^2 = \frac{1}{2\pi e}e^{2H(Z)}. \tag{3}$$

Thus, the entropy power inequality is an inequality between effective variances, where effective variance (entropy power) is simply the variance of the normal random variable with the same entropy.

Manuscript received October 21, 1983; revised July 16, 1984. This work was partially supported by the National Science Foundation under Grant ECS82-11568, the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), the Instituto de Pesquisas Espaciais (INPE) of Brazil, and the Joint Services Electronics Program under Grant DAAG-29-81-K-0057. This work was presented at the IEEE International Symposium on Information Theory, St. Jovite, PQ, Canada, September 26–30, 1983.
T. Cover is with the Department of Electrical Engineering and Statistics, Stanford University, Stanford, CA 94305 USA.
M. Costa is with Instituto de Pesquisas Espaciais, Caixa Postal 515, São José dos Campos, SP 12200, Brazil.

The preceeding equations allow the entropy power inequality to be rewritten in the equivalent form

$$H(X+Y) \geq H(X'+Y') \tag{4}$$

where X' and Y' are independent normal variables with corresponding entropies $H(X') = H(X)$ and $H(Y') = H(Y)$. Verification of this restatement follows from the use of (1) to show that

$$\begin{aligned}\frac{1}{2\pi e}e^{2H(X+Y)} &\geq \frac{1}{2\pi e}e^{2H(X)} + \frac{1}{2\pi e}e^{2H(Y)} \\ &= \frac{1}{2\pi e}e^{2H(X')} + \frac{1}{2\pi e}e^{2H(Y')} \\ &= \sigma^2_{X'} + \sigma^2_{Y'} \\ &= \sigma^2_{X'+Y'} \\ &= \frac{1}{2\pi e}e^{2H(X'+Y')}\end{aligned} \tag{5}$$

where the penultimate equality follows from the fact that the sum of two independent normals is normal with a variance equal to the sum of the variances.

By the same line of reasoning, the entropy power inequality for independent random n-vectors $\mathbf{X}$ and $\mathbf{Y}$ that is given by

$$e^{2H(\mathbf{X}+\mathbf{Y})/n} \geq e^{2H(\mathbf{X})/n} + e^{2H(\mathbf{Y})/n} \tag{6}$$

can be recast as

$$H(\mathbf{X}+\mathbf{Y}) \geq H(\mathbf{X}'+\mathbf{Y}') \tag{7}$$

where $\mathbf{X}', \mathbf{Y}'$ are independent multivariate normal random vectors with proportional covariance matrices and corresponding entropies.

II. The Brunn–Minkowski Inequality

Let A and B be two measurable sets in $\mathbf{R}^n$. The set sum $C = A + B$ of these sets may be written as

$$C = \{x + y : x \in A,\ y \in B\}. \tag{8}$$

Let $V(A)$ denote the volume of A. The Brunn–Minkowski inequality [2], [3] states that

$$V^{1/n}(A+B) \geq V^{1/n}(A) + V^{1/n}(B). \tag{9}$$

To recast this inequality, we observe that an n sphere S with radius r has volume

$$V(S) = c_n r^n. \tag{10}$$

Thus if S is a sphere with volume V, its radius is

$$r = (V/c_n)^{1/n}. \tag{11}$$

Hence the Brunn–Minkowski inequality can be viewed as an inequality between the radii of the spherical equivalents of the sets.

Rewriting the Brunn–Minkowski inequality following the model in (5) gives

$$V(A+B) \geq V(A'+B') \tag{12}$$

where A' and B' are spheres with volumes $V(A') = V(A)$ and $V(B') = V(B)$.

0018-9448/84/1100-0837$01.00

III. Comparisons

Inspecting the convolution reveals further similarity between the entropy power and Brunn–Minkowski inequalities. The entropy $H(X+Y)$ is a functional of the convolution, i.e.,

$$H(X+Y) = -\int f_Z \ln f_Z, \tag{13}$$

where

$$f_Z(z) = (f_X * f_Y)(z) = \int f_X(t) f_Y(z-t)\, dt.$$

Similarly, the volume $V(A+B)$ is a functional of a convolution, i.e.,

$$V(A+B) = \int I_C(z)\, dz \tag{14}$$

where $I_C(z)$ is the indicator function for $C = A + B$ given by

$$I_C(z) = \max_t I_A(t) I_B(z-t). \tag{15}$$

We note that $f_Z(z)$ is the L_1 norm of $h_z(t) = f_X(t) f_Y(z-t)$; $I_C(z)$ is the L_∞ norm of $h_z(t) = I_A(t) I_B(z-t)$; and L_1 and L_∞ are dual spaces.

Finally e^H, like V, is a measure of volume. For example, for all random variables X with support set A, we have $H(X) \le \ln V(A)$ with equality if the probability density is uniform over A. Moreover, from the Asymptotic Equipartition Property, we know that the volume of the set of ϵ-typical n-sequences $(X_1, X_2, \cdots, X_n)$, with X_i independent and identically distributed (i.i.d.) according to $f(x)$, is equal to $e^{nH(X)}$ to the first order in the exponent. Thus, $e^{H(\mathbf{X})} = e^{nH(X)}$ is the volume of the typical set for $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. To suggest a link between the Gaussian distribution and spheres, we note that the ϵ-typical set of n sequences $(X_1, X_2, \cdots, X_n)$ is given by a sphere when X_i are i.i.d. according to the Gaussian distribution.

These observations suggest not only that the two inequalities may be different manifestations of the same underlying idea, but that there also may be a continuum of inequalities between L_1 and L_∞ with their respective natural definitions of volume.

In spite of the obvious similarity between the above inequalities, there is no apparent similarity between any of the known proofs of the Brunn–Minkowski inequality [2]–[4] and the Stam and Blachman [5], [6] proofs of the entropy power inequality, nor have we succeeded in finding a new common proof. Nevertheless, the similarity of these inequalities suggests that we may find new results relating to entropies from known results in geometry and vice versa. We present two applications of this reasoning.

IV. Isoperimetric Inequalities

It is known that the sphere minimizes surface area for given volume. A proof follows immediately from the Brunn–Minkowski inequality [4] as shown below. For "regular" sets A, the surface area $S(A)$ of A is given by

$$S(A) = \lim_{\epsilon \to 0} \frac{V(A + S_\epsilon) - V(A)}{\epsilon} \tag{16}$$

where S_ϵ is a sphere of radius $\epsilon > 0$. Using the Brunn–Minkowski inequality (12) we have

$$S(A) \ge \lim_{\epsilon \to 0} \frac{V(A' + S_\epsilon) - V(A')}{\epsilon} = S(A') \tag{17}$$

where A' is a sphere with volume $V(A') = V(A)$. Thus the surface area of A is greater than that of a sphere with the same volume.

We now proceed to perform the same steps on the entropy power. Let $\mathbf{X}$ have a suitably smooth density $p(\mathbf{x})$. We define $S(\mathbf{X})$, the "surface area" of a multivariate random variable $\mathbf{X}$, as

$$S(\mathbf{X}) \triangleq \lim_{\epsilon \to 0} \frac{e^{2H(\mathbf{X}+\mathbf{Z}_\epsilon)} - e^{2H(\mathbf{X})}}{\epsilon} \tag{18}$$

where $\mathbf{Z}_\epsilon$ is Gaussian with covariance matrix ϵI and $\mathbf{Z}_\epsilon$ is independent of $\mathbf{X}$. Thus $S(\mathbf{X})$ is the rate of change of e^{2H} when a small normal random variable is added. We may write

$$S(\mathbf{X}) = \lim_{t_0 \to 0} \frac{d}{dt} e^{2H(\mathbf{X}+\mathbf{Z}_t)} \Big|_{t=t_0}. \tag{19}$$

Now let $\mathbf{X}_t \triangleq \mathbf{X} + \mathbf{Z}_t$ have density denoted by $p_t(x_t)$. Assuming $H(\mathbf{X})$ is finite, we prove in the Appendix that

$$\frac{d}{dt} H(\mathbf{X}_t) = \frac{1}{2} E \frac{\|\nabla p_t\|^2}{p_t^2}$$

where $\|\nabla p_t\|$ denotes the norm of the gradient of $p_t(x_t)$. It follows that

$$S(\mathbf{X}) = E \frac{\|\nabla p\|^2}{p^2} e^{2H(\mathbf{X})}. \tag{20}$$

From the entropy power inequality (7), we have

$$e^{2H(\mathbf{X}+\mathbf{Z}_\epsilon)} \ge e^{2H(\mathbf{X}'+\mathbf{Z}_\epsilon)} \tag{21}$$

where $\mathbf{X}'$ is a Gaussian n vector with covariance matrix of the form $\sigma^2 I$ such that $H(\mathbf{X}') = H(\mathbf{X})$. Thus, from (18) and (21), we obtain the following bound on $S(\mathbf{X})$:

$$S(\mathbf{X}) \ge \lim_{\epsilon \to 0} \frac{e^{2H(\mathbf{X}'+\mathbf{Z}_\epsilon)} - e^{2H(\mathbf{X}')}}{\epsilon} = S(\mathbf{X}'). \tag{22}$$

For the spherical Gaussian vector $\mathbf{X}'$, the "surface area" is found from (20) to be

$$S(\mathbf{X}') = \frac{2\pi e n e^{2H(\mathbf{X}')}}{e^{2H(\mathbf{X}')/n}}. \tag{23}$$

Recalling that $H(\mathbf{X}') = H(\mathbf{X})$, we combine (20), (22), and (23) to obtain the entropy analog of the isoperimetric inequality

$$\frac{1}{n} E \frac{\|\nabla p\|^2}{p^2} \ge 2\pi e e^{-2H(\mathbf{X})/n}. \tag{24}$$

Let $J(\mathbf{X})$ denote the trace of the Fisher information matrix for the translation family of densities $\{p(\mathbf{x} - \boldsymbol{\theta})\}$, $\boldsymbol{\theta} \in \mathbf{R}^n$, and $\mathbf{x} \in \mathbf{R}^n$. We formalize the above inequality in a theorem.

Theorem: Suppose the multivariate random variable $\mathbf{X}$ has finite entropy $H(\mathbf{X})$, and suppose $J(\mathbf{X})$ exists. Let $\mathbf{X}_t$ denote $\mathbf{X} + \mathbf{Z}_t$, where $\mathbf{Z}_t$ is spherical normal with covariance matrix tI and independent of $\mathbf{X}$. If $\mathbf{X}$ has a continuous density $p(\mathbf{x})$ so that

$$J(\mathbf{X}) = E \frac{\|\nabla p\|^2}{p^2}$$

and if the integrals $H(\mathbf{X}_t)$ and $J(\mathbf{X}_t)$ converge uniformly near $t = 0$ so that

$$H(\mathbf{X}) = \lim_{t \to 0} H(\mathbf{X}_t)$$

and

$$J(\mathbf{X}) = \lim_{t \to 0} J(\mathbf{X}_t)$$

then

$$J(\mathbf{X}) \ge J(\mathbf{X}')$$

where $\mathbf{X}'$ is spherical normal with entropy $H(\mathbf{X}') = H(\mathbf{X})$.

This establishes that the Gaussian distribution minimizes the trace of the Fisher information matrix given an entropy constraint. The scalar version of the above relation was proved in [5].

V. Concavity

Here we present an example of how a result involving entropy may yield a conjecture in geometry. Let $\boldsymbol{X}$ be an arbitrary random vector, let $\boldsymbol{Z}_t$ be multivariate Gaussian with covariance matrix tI, and let $\boldsymbol{Z}_t$ be independent of $\boldsymbol{X}$. In [7] and [8] the entropy power of $\boldsymbol{X} + \boldsymbol{Z}_t$, given by

$$\frac{e^{2H(X+Z_t)/n}}{2\pi e},$$

is shown to be a concave function of the added Gaussian noise variance t. This suggests the following conjecture.

Conjecture: Let A be an arbitrary measurable set, and let S be a sphere of unit radius in $\boldsymbol{R}^n$. Then $V^{1/n}(A + tS)$ is a concave function of $t \geq 0$.

We prove that the conjecture is true if A is convex.

Theorem: If A is convex and S is a sphere, then $V^{1/n}(A + tS)$ is a concave function of $t \geq 0$.

Proof: Let $\theta \in [0,1]$. Then we have

$$\begin{aligned} V^{1/n}(A + \theta S) &= V^{1/n}((1-\theta)A + \theta A + \theta S) \\ &\geq V^{1/n}((1-\theta)A) + V^{1/n}(\theta A + \theta S) \\ &= (1-\theta)V^{1/n}(A) + \theta V^{1/n}(A + S) \end{aligned} \quad (25)$$

where the convexity of the set A in the first line and the Brunn–Minkowski inequality in the second line was used. Since A is an arbitrary convex set, scaling $A + tS$ completes the proof.

As far as we know, the conjecture is unresolved for arbitrary measurable sets in $\boldsymbol{R}^n$.

VI. Concluding Remarks

We still do not know if the Brunn–Minkowski and entropy power inequalities have a common underlying idea leading to similar proofs. The inequalities look the same ((7) and (12)), conditions for equality are similar, and both involve measures of volume.

Jânos Körner calls our attention to some related problems in combinatorics where the Brunn–Minkowski inequality and the isoperimetric property are applied to Hamming spaces. An account of these problems is given in [10].

Acknowledgment

The authors wish to thank Brad Efron, Paul Shields, and Rudolf Ahlswede for their discussions. Thanks are also due to the reviewers for their useful comments.

Appendix

Let $\boldsymbol{X}$ be a vector-valued random variable having arbitrary density $p(\boldsymbol{x})$ and finite entropy $H(\boldsymbol{X})$. Let $\boldsymbol{X}_t = \boldsymbol{X} + \boldsymbol{Z}_t$ be the sum of $\boldsymbol{X}$ and an independent spherical multivariate normal random variable $\boldsymbol{Z}_t$ with covariance tI. Then $\boldsymbol{X}_t$ has density $p_t(\boldsymbol{x}_t)$ given by

$$p_t(\boldsymbol{x}_t) = \frac{1}{(2\pi t)^{n/2}} \int_{\boldsymbol{R}^n} p(\boldsymbol{x}) \exp\left(\frac{-\|\boldsymbol{x}_t - \boldsymbol{x}\|^2}{2t}\right) d\boldsymbol{x}. \quad (A1)$$

Due to the smoothing properties of the normal distribution, we can differentiate the above expression inside the integral (the integrand is continuous and differentiable in t) to show that $p_t(\boldsymbol{x}_t)$ satisfies the diffusion (heat) equation

$$\frac{d}{dt} p_t(\boldsymbol{x}_t) = \frac{1}{2} \nabla^2 p_t(\boldsymbol{x}_t) \quad (A2)$$

where

$$\nabla^2 p_t(\boldsymbol{x}_t) = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_{(i)t}^2} p_t(\boldsymbol{x}_t). \quad (A3)$$

Interchanging derivative and integral once more gives

$$\begin{aligned} \frac{d}{dt} H(\boldsymbol{X}_t) &= -\int_{\boldsymbol{R}^n} \frac{d}{dt} p_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t - \int_{\boldsymbol{R}^n} \left(\frac{d}{dt} p_t(\boldsymbol{x}_t)\right) \log p_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t \\ &= 0 - \frac{1}{2} \int_{\boldsymbol{R}^n} (\nabla^2 p_t(\boldsymbol{x}_t)) \log p_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t. \end{aligned} \quad (A4)$$

We now recall Green's identity [9]: if $\phi(\boldsymbol{x})$ and $\psi(\boldsymbol{x})$ are twice continuously differentiable functions in $\boldsymbol{R}^n$ and if V is any set bounded by a piecewise smooth, closed, and oriented surface S in $\boldsymbol{R}^n$, then

$$\int_V \phi \nabla^2 \psi \, dV = \int_S \phi \nabla \psi \cdot \boldsymbol{ds} - \int_V \nabla \phi \cdot \nabla \psi \, dV \quad (A5)$$

where $\nabla \psi$ denotes the gradient of ψ, $\boldsymbol{ds}$ denotes the elementary area vector, and $\nabla \psi \cdot \boldsymbol{ds}$ is the inner product of these two vectors. This identity plays the role of integration by parts in $\boldsymbol{R}^n$.

To apply Green's identity to (A4), we let V_r be the n sphere of radius r centered at the origin and having surface S_r. Then we use Green's identity on V_r and S_r with $\phi(\boldsymbol{x}_t) = \log p_t(\boldsymbol{x}_t)$ and $\psi(\boldsymbol{x}_t) = p_t(\boldsymbol{x}_t)$ and take the limit as $r \to \infty$. We can show that the surface integral over S_r vanishes in the limit [7], [8] provided that $H(\boldsymbol{X}_t)$ is finite. Hence we obtain

$$\begin{aligned} \frac{d}{dt} H(\boldsymbol{X}_t) &= -\frac{1}{2} \int_{\boldsymbol{R}^n} \nabla p_t(\boldsymbol{x}_t) \cdot \nabla \log p_t(\boldsymbol{x}_t)\, d\boldsymbol{x}_t \\ &= \frac{1}{2} \int_{\boldsymbol{R}^n} \frac{\|\nabla p_t(\boldsymbol{x}_t)\|^2}{p_t(\boldsymbol{x}_t)}\, d\boldsymbol{x}_t \end{aligned} \quad (A6)$$

as desired.

Equation (A6) can also be written as

$$\frac{d}{dt} H(\boldsymbol{X}_t) = \frac{1}{2} E \frac{\|\nabla p_t\|^2}{p_t^2}. \quad (A7)$$

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* vol. 27, pp. 623–656, Oct. 1948.
[2] B. Grunbaum, *Convex Polytopes.* London: Interscience, 1967.
[3] H. G. Eggleston, *Convexity.* Cambridge: University Press, 1958.
[4] H. Hadwiger and D. Ohmann, "Brunn-Minkowskischer Satz und Isoperimetrie," *Math. Z.*, vol. 66, pp. 1–8, 1956.
[5] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. and Contr.*, vol. 2, pp. 101–112, June 1959.
[6] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267–271, Apr. 1965.
[7] M. H. M. Costa, "Interference channels," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Feb. 1983.
[8] M. H. M. Costa, "A new entropy power inequality," *IEEE Trans. Inform. Theory*, to be published.
[9] J. Amazigo and L. Rubenfeld, *Advanced Calculus and its Applications to the Engineering and Physical Sciences.* New York: Wiley, 1980.
[10] R. Ahlswede and G. O. H. Katona, "Contributions to the geometry of Hamming spaces," *Discrete Mathematics*, vol. 17, pp. 1–22, Jan. 1977.