Information Theoretic Inequalities for Contoured Probability Distributions

Onur G. Guleryuz , Erwin Lutwak, Deane Yang, Gaoyong Zhang

O. G. Guleryuz (onur@vision.poly.edu) is with the Department of Electrical Engineering, and E. Lutwak (elutwak@poly.edu), D. Yang (dyang@poly.edu), and G. Zhang (gzhang@poly.edu) are with the Department of Mathematics, Polytechnic University, Brooklyn, New York.

Abstract

We show that for a special class of probability distributions that we call *contoured distributions*, information theoretic invariants and inequalities are equivalent to geometric invariants and inequalities of bodies in Euclidean space associated with the distributions. Using this, we obtain characterizations of contoured distributions with extremal Shannon and Renyi entropy. We also obtain a new reverse information theoretic inequality for contoured distributions.

Keywords

Brunn-Minkowski, convex bodies, elliptically contoured, entropy, Fisher information, inequalities, isoperimetric inequalities

I. Introduction

That there is a connection between information theory and geometry was demonstrated by Lieb [1] and later, unaware of Lieb's earlier work, proposed by Costa and Cover [2]. This connection was developed in detail by Cover, Dembo, and Thomas [3] (also, see [4]). We show in this paper that an even closer connection can be made through the study of a special class of probability distributions that we call *contoured distributions.* Known geometric inequalities can be obtained by applying basic information theoretic inequalities to contoured distributions. Conversely, we show that new reverse information theoretic inequalities for convex contoured distributions can be obtained from recently established reverse geometric inequalities for convex bodies. An open question is whether these reverse inequalities hold for a larger class of probability distributions.

Our work was originally motivated by ideas introduced by Lieb [1] and Costa and Cover [2] and developed further by Cover, Dembo, and Thomas [3]. Costa and Cover observed that the Brunn–Minkowski inequality, which is about the geometry of bodies, closely resembles the Shannon power inequality, which is an information theoretic inequality for probability distributions. They speculated that the similarity was not coincidental. It turned out that Lieb [1] had already shown that the Brunn-Minkowski and Shannon inequalities both followed from the sharp Young's inequality proved by Beckner [5] (also, see Brascamp-Lieb [6]). Cover, Dembo, and Thomas [3] found other parallels between information theory and geometry, showing, among other things, that the isoperimetric inequality in geometry is analogous to an inequality relating the Fisher information and entropy of a probability distribution.

The results obtained by Cover, Dembo, and Thomas apply to any probability distribution that satisfies mild regularity assumptions. They also show that inequalities involving determinants of positive definite symmetric matrices can be proved by applying information theoretic inequalities to Gaussian distributions. We show here that this approach can be generalized and that geometric inequalities for bodies in Euclidean space can be obtained by applying information theoretic inequalities to the larger class of contoured distributions.

Roughly speaking, a distribution is called *contoured*, if there is a set in $\mathbf{R}^n$ such that any level set of the probability density function is a dilate of this set. We call this set the *contour body* of the distribution. Any reasonable star-shaped set in $\mathbf{R}^n$ that contains the origin in its interior can be realized as the the contour body of a contoured distribution. See §II for precise statements. Contoured distributions whose contour body is an ellipsoid centered at the origin are known as elliptically contoured distributions and have been studied extensively (see, for example, [7]).

The probability distribution function of a contoured distribution can always be represented as the composition of two functions, one that depends only on the norm of the random vector and the other that depends only on the direction. Using this observation along with polar coordinates on $\mathbf{R}^n$, any information theoretic invariant of the distribution that is defined by an integral can be broken up into two parts, a radial piece and an angular piece. A closer examination of the angular part shows that it is equal to a geometric invariant of the contour body. This approach to studying natural integrals involving contoured functions is not new; it has been, for example, also used by Yamada, Tazaki, and Gray [8] to obtain formulas and results for norm–based distortion functions that are similar to those presented here.[1]

We provide a precise definition of a contoured probability distribution and its associated contour body in §II. Next, in §III we use integration in polar coordinates to obtain formulas for the mean and entropy of a contoured distribution in terms of the center of mass and

[1]We would like to thank the referee for bringing this paper to our attention.

volume of the contour body. We also show that the covariance and Fisher information matrices of a contoured distribution are closely related to geometrically natural ellipsoids associated with the contour body.

In §IV we recall some well-known inequalities relating the entropy of a probability distribution to its covariance and Fisher information. These results, along with the polar decomposition of a contoured distribution, lead to characterizations of a contoured distribution with extremal entropy in terms of its radial profile and contour body (see §II for the definitions). These characterizations, in turn, imply two geometric inequalities, one classical and one recently discovered.

This naturally leads to the question of whether new information theoretic inequalities can be proved using geometric inequalities. We demonstrate one case of this, where a reverse geometric inequality recently established by the authors implies a new reverse information theoretic inequality for contoured probability distributions. We end the paper with an important conjectured geometric inequality and its analogue in information theory.

II. Definitions

A. Preliminaries

Given an integer $n \geq 1$, we shall denote n–dimensional Euclidean space by $\mathbf{R}^n$. The standard Euclidean norm on $\mathbf{R}^n$ will be denoted $\|x\|$, for any $x \in \mathbf{R}^n$.

The sphere of unit vectors in $\mathbf{R}^n$ will be denoted S. Recall that polar coordinates on $\mathbf{R}^n$ are given by the map

$$\begin{aligned} S \times [0,\infty) &\to \mathbf{R}^n; \\ (u,r) &\mapsto x = ru, \end{aligned}$$

where r and u can be recovered from $x \in \mathbf{R}^n$ using the formulas

$$r = \|x\|, \; u = \frac{x}{\|x\|}.$$

Standard Lebesgue measure on $\mathbf{R}^n$ will be denoted dx. The standard surface area measure on S will be denoted du. Using polar coordinates, Lebesgue measure can be written as

$$dx = r^{n-1}\, dr\, du.$$

Definition: We shall denote *the ball of vectors* with Euclidean norm at most 1 by B and *its volume* by v_n; note that

$$v_n = \int_B dx = \frac{1}{n} \int_S du. \tag{1}$$

B. Definition of a contoured distribution

Definition: A probability density function $f(x)$ defines a *contoured distribution*, if using polar coordinates there exists a decomposition of the probability density function f of the following form:

$$f(x) = c\phi(r\lambda(u)), \tag{2}$$

where

$$\begin{aligned} c &= \text{ a positive constant;} \\ \lambda &: S \to (0, \infty); \\ \phi &: [0, \infty) \to [0, \infty). \end{aligned}$$

The constant c can, of course, be absorbed into the function ϕ, but it will be convenient to have the extra degree of freedom in the normalizations discussed later. Given such a decomposition, we call the function ϕ a *radial profile* and the set

$$C = \{ru \in \mathbf{R}^n \mid r\lambda(u) \leq 1\} \tag{3}$$

a *contour body* of the distribution with density function $f(x)$.

It is convenient to extend λ to be a homogeneous function on $\mathbf{R}^n$ by setting $\lambda(0) = 0$ and for $x \neq 0$,

$$\lambda(x) = \|x\| \lambda \left(\frac{x}{\|x\|} \right).$$

Using this definition, λ is positive away from 0 and *homogeneous,* which means that for any $x \in \mathbf{R}^n$ and $t \in \mathbf{R}, \ t > 0$,

$$\lambda(tx) = t\lambda(x).$$

With this definition of λ, the decomposition in (2) can be written as

$$f(x) = c\phi(\lambda(x)).$$

C. The contour body and its shape function

Definition: We will call any function $\lambda : \mathbf{R}^n \to \mathbf{R}$ that is positive away from 0 and homogeneous a *shape function*.

Definition: A set $C \subset \mathbf{R}^n$ is said to be *star–shaped*, if for any $x \in C$, the line segment joining 0 to x lies in C.

Definition: We will call any compact star–shaped set that contains a neighborhood of 0 in $\mathbf{R}^n$ a *contour body*.

Associated to any shape function λ is the contour body

$$C_\lambda = \{x \in \mathbf{R}^n \mid \lambda(x) \leq 1\}. \tag{4}$$

Conversely, associated to any contour body C is the shape function defined by

$$\lambda_C(x) = \inf \left\{ t > 0 \mid \frac{x}{t} \in C \right\}, \tag{5}$$

for any $x \in \mathbf{R}^n$.

Lemma 1: Equations (4) and (5) define a one–to–one correspondence between contour bodies and shape functions.

For example, if the contour body C is a unit ball centered at the origin, then $\lambda(x) = \|x\|$.

For convenience we will restrict our attention to contour bodies with continuous shape functions. Stronger assumptions on the contour body will be stated explicitly as needed.

The most basic invariant of C is its volume $V(C)$, and its formula in terms of the shape function λ_C is needed throughout. Using polar coordinates, it is given by

$$V(C) \;\; = \;\; \frac{1}{n} \int_S \lambda_C(u)^{-n} \, du, \tag{6}$$

where du is the standard surface area measure on the unit sphere S.

D. The radial profile

Let f be the density function of a contoured distribution given by $f(x) = c\phi(\lambda(x))$. Since the total integral of f is equal to 1,

$$\begin{aligned} 1 &= \int_{\mathbf{R}^n} f(x)\,dx \\ &= \int_S \int_0^\infty c\phi(r\lambda(u))\, r^{n-1}\,dr\,du \\ &= \int_S \int_0^\infty c\phi(s)(\lambda(u)^{-1}s)^{n-1}\lambda(u)^{-1}\,ds\,du \\ &= c\left(\frac{1}{n}\int_S \lambda(u)^{-n}\,du\right)\left(n\int_0^\infty \phi(s)s^{n-1}\,ds\right) \\ &= cV(C)\left(n\int_0^\infty \phi(s)s^{n-1}\,ds\right) \qquad (7) \end{aligned}$$

where $V(C)$ is given by (6). It follows that a radial profile must be a measurable function $\phi : [0,\infty) \to [0,\infty)$ such that

$$\int_0^\infty \phi(r)r^{n-1}\,dr < \infty. \qquad (8)$$

Note that the radial profile can even be a generalized function, such as a Dirac delta function, as long as (8) holds. Stronger assumptions about the radial profile ϕ will be stated explicitly when needed.

E. Normalization of c, ϕ, and C

First, observe that by (7) the constant c is uniquely determined by the radial profile ϕ and the contour body C. Second, the decomposition $f(x) = c\phi(\lambda_C(x))$ of the density function f of a contoured distribution is not unique. Given positive constants a and b, if

$$\begin{aligned} \widetilde{C} &= b^{-1}C = \left\{\frac{x}{b} \mid x \in C\right\}, \\ \lambda_{\widetilde{C}}(x) &= b\lambda_C(x), \\ \tilde{\phi}(r) &= a\phi(b^{-1}r), \\ \tilde{c} &= a^{-1}c, \end{aligned}$$

it is clear that $f(x) = \tilde{c}\tilde{\phi}(\lambda_{\widetilde{C}}(x))$.

The radial profile ϕ can be partially normalized by requiring the spherically symmetric function

$$f_\phi(x) = \phi(\|x\|) \qquad (9)$$

to be a probability density function, so that

$$1 = \int_{\mathbf{R}^n} \phi(\|x\|)\, dx = n v_n \int_0^\infty \phi(r) r^{n-1}\, dr,$$

where v_n is given by (1). If this holds, then by (7)

$$f(x) = \frac{v_n}{V(C)} \phi(\lambda(x)) \tag{10}$$

Throughout this paper, we will always use a decomposition of f that satisfies (10).

However, this normalization still does not uniquely determine ϕ, C, and c, because we can still rescale as follows:

$$\begin{aligned} \tilde{\lambda}(x) &= b\lambda(x) \\ \tilde{\phi}(r) &= b^{-n}\phi(b^{-1}r) \\ \tilde{c} &= b^n c \end{aligned}$$

Ideally, there should be a natural unique normalized decomposition of f. Three possibilities are described in the appendix. We do not find any of them satisfactory, and our results in the following sections are presented without using any of the normalizations discussed in the appendix.

III. Invariants of contoured distributions

In this section we show that by using polar coordinates to integrate over $\mathbf{R}^n$, information theoretic invariants of a contoured distribution always split nicely into two factors, one that depends only on the radial profile and the other only on the contour body. In particular, we observe that for a contoured distribution with zero mean the covariance and Fisher information matrices are closely related to naturally defined ellipsoids associated with the contour body. Here, by the Fisher information matrix we mean the $n \times n$ Fisher information matrix obtained by translating the probability distribution in $\mathbf{R}^n$.

A. The mean

Recall that a body $C \subset \mathbf{R}^n$ with constant mass density has a *center of mass*

$$\hat{x} = \frac{1}{V(C)} \int_C x\, dx. \tag{11}$$

Lemma 2: Given a contoured distribution on $\mathbf{R}^n$ with probability density function f given by (10), the mean of the distribution is given by

$$\overline{x} = \frac{n+1}{n}\hat{r}\hat{x},$$

where $\hat{x}$ is the center of mass of C and

$$\hat{r} \;=\; \int_{\mathbf{R}^n} \|x\|\phi(\|x\|)\,dx.$$

Proof: Using integration in polar coordinates,

$$\begin{aligned}
\overline{x} &= \int_{\mathbf{R}^n} xf(x)\,dx \\
&= \frac{v_n}{V(C)}\int_S\int_0^\infty ru\phi(r\lambda(u))r^{n-1}\,dr\,du \\
&= \left(v_n\int_0^\infty s^n\phi(s)\,ds\right)\frac{1}{V(C)}\int_S u\lambda(u)^{-n-1}\,du
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\hat{x} &= \frac{1}{V(C)}\int_C x\,dx \\
&= \frac{1}{V(C)}\int_S\int_0^{(\lambda(u))^{-1}} r^n u\,dr\,du \\
&= \frac{1}{(n+1)V(C)}\int_S u\lambda(u)^{-n-1}\,du
\end{aligned}$$

and

$$\begin{aligned}
\hat{r} &= \int_{\mathbf{R}^n}\|x\|\phi(\|x\|)\,dx \\
&= \int_S\int_0^\infty r^n\phi(r)\,dr\,du \\
&= nv_n\int_0^\infty r^n\phi(r)\,dr
\end{aligned}$$

■

It follows that a contoured distribution with bounded mean has mean equal to $0 \in \mathbf{R}^n$ if and only if the center of mass of its contour body is at the origin.

Definition: A contoured distribution is *balanced* if the center of mass of its contour body is at the origin.

It is worth noting that if C is an ellipsoid centered at the origin or any other origin–symmetric body ($\lambda_C(-x) = \lambda_C(x)$), its center of mass is at the origin and therefore the contoured distribution is balanced.

B. Volume and entropy

Definition: The (differential) *Shannon entropy* of a probability density function f on $\mathbf{R}^n$ is defined by

$$h[f] = -\int_{\mathbf{R}^n} f(x) \log f(x)\, dx.$$

More generally, given $q \in \mathbf{R}$ such that $q > 0$ and $q \neq 1$, the q–*Renyi entropy* is defined by

$$h_q[f] = -\frac{1}{q-1} \log \int_{\mathbf{R}^n} f(x)^q \, dx.$$

Note that $\lim_{q \to 1} h_q[f] = h[f]$ and therefore we can define $h_1 = h$. Furthermore, if f has compact support, then the 0–*Renyi entropy* $h_0[f]$ is well–defined and equal to

$$h_0[f] = \log V(\operatorname{supp} f),$$

where

$$\operatorname{supp} f = \{x \mid f(x) > 0\}.$$

Lemma 3: Given $q \geq 0$ and a contoured probability distribution with density function f given by (10), the q–Renyi entropy of f is related to the q–Renyi entropy of f_ϕ, defined by (9) and the volume of the contour body C via:

$$h_q[f] \;=\; h_q[f_\phi] + \log \frac{V(C)}{v_n}.$$

Proof: Note that

$$\begin{aligned}
\int_{\mathbf{R}^n} f(x)^q \, dx &= \left(\frac{v_n}{V(C)}\right)^q \int_S \int_0^\infty \phi(r\lambda(u))^q r^{n-1} \, dr \, du \\
&= \left(\frac{v_n}{V(C)}\right)^q \int_0^\infty \phi(s)^q s^{n-1} \, ds \int_S \lambda(u)^{-n} \, du \\
&= \left(\frac{v_n}{V(C)}\right)^{q-1} v_n n \int_0^\infty \phi(s)^q s^{n-1} \, ds \\
&= \left(\frac{v_n}{V(C)}\right)^{q-1} \int f_\phi^q(x) \, dx,
\end{aligned}$$

where we have used (10), (9), and (1). The formula for $h[f]$ can be obtained either by taking the limit $q \to 1$ of the formula for $h_q[f]$ or by a direct computation similar to the one above. ∎

C. Ellipsoids

Two of the most important invariants associated with a probability distribution are its covariance matrix and its Fisher information matrix, which are positive definite symmetric matrices and therefore define ellipsoids in $\mathbf{R}^n$. A contour body in $\mathbf{R}^n$ also has two naturally defined ellipsoids associated with it. We show that for a balanced contoured distribution, the ellipsoids associated with the contour body coincide with the ellipsoids associated with the distribution.

First, recall that a symmetric positive definite $n \times n$ matrix A uniquely determines an ellipsoid E in $\mathbf{R}^n$ centered at the origin:

$$E = \{x \in \mathbf{R}^n \mid x \cdot A^{-1} x \leq 1\}. \tag{12}$$

Conversely, given an ellipsoid E in $\mathbf{R}^n$ centered at the origin, there is a uniquely determined symmetric positive definite $n \times n$ matrix A such that (12) holds.

Definition: Given a symmetric positive definite matrix A, we will denote *the corresponding ellipsoid* by E_A. Conversely, given an ellipsoid $E \subset \mathbf{R}^n$ centered at the origin, we will denote *the corresponding symmetric positive definite matrix* by $[E]$.

We also recall that the volume of an ellipsoid is given by

$$V(E_A) = \sqrt{\det A}\, v_n, \tag{13}$$

where v_n is given by (1).

D. Two ellipsoids associated with a set in $\mathbf{R}^n$

Definition: Given a measurable set $K \subset \mathbf{R}^n$ with positive volume and center of mass located at the origin, consider the following positive definite symmetric matrix

$$A_{ij} = \frac{n+2}{V(K)} \int_K x_i x_j \, dx, \tag{14}$$

and denote the corresponding ellipsoid by

$$\Gamma_2 K = E_A. \tag{15}$$

The ellipsoid is normalized so that for the unit ball $B \subset \mathbf{R}^n$, $\Gamma_2 B = B$. The ellipsoid $\Gamma_2 K$ is called the *Legendre ellipsoid* or the *ellipsoid of inertia* in physics (see [9] for a discussion of its history). The matrix $[\Gamma_2 K]$ is, up to a scale factor, the covariance matrix of the uniform distribution on K.

More recently, Lutwak, Yang, and Zhang [10] have introduced another ellipsoid that is naturally associated with a bounded convex set that contains the origin in its interior. It is in some sense dual to the Legendre ellipsoid and can be described as follows.

Definition: Given a convex contour body $K \subset \mathbf{R}^n$ with shape function λ, let F be the positive definite symmetric $n \times n$ matrix such that for any vector $v \in R^n$,

$$v \cdot Fv = \frac{n}{V(K)} \int_K (v \cdot \nabla\lambda)^2 \, dx \tag{16}$$

where ∇ denotes gradient, and denote

$$\Gamma_{-2} K = E_{F^{-1}}. \tag{17}$$

Note that $\Gamma_{-2} K$ is defined using F^{-1}. Again, the definition is normalized so that $\Gamma_{-2} B = B$ for the unit ball B.

E. Two ellipsoids associated with a probability distribution

Definition: The *covariance matrix* of a probability distribution with density function f on $\mathbf{R}^n$ with finite variance and zero mean, is given by

$$A_{ij} = \int_{\mathbf{R}^n} x_i x_j f(x) \, dx. \tag{18}$$

Denote the associated ellipsoid by

$$\Gamma_2 f = E_A. \tag{19}$$

Lemma 4: Given a contoured distribution having density f given by (10) with finite variance and zero mean,

$$\Gamma_2 f = \hat{\sigma} \Gamma_2 C,$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \int_{\mathbf{R}^n} \|x\|^2 \phi(\|x\|) \, dx. \tag{20}$$

In particular, if the decomposition (10) is normalized as in Lemma 17, then the two ellipsoids $\Gamma_2 f$ and $\Gamma_2 C$ coincide.

Proof: First, observe that using polar coordinates on (14), given any $v \in \mathbf{R}^n$,

$$\begin{aligned} v \cdot [\Gamma_2 C]v &= \frac{n+2}{V(C)} \int_C (v \cdot x)^2 \, dx \\ &= \frac{n+2}{V(C)} \int_S \int_0^{(\lambda(u))^{-1}} (v \cdot u)^2 r^{n+1} dr \, du \\ &= \frac{1}{V(C)} \int_S (v \cdot u)^2 \lambda(u)^{-n-2} \, du. \end{aligned}$$

Therefore,

$$\begin{aligned} v \cdot [\Gamma_2 f]v &= \int_{\mathbf{R}^n} (v \cdot x)^2 f(x) \, dx \\ &= \frac{v_n}{V(C)} \int_S \int_0^\infty r^{n+1} \phi(r\lambda(u))(v \cdot u)^2 \, dr \, du \\ &= \frac{v_n}{V(C)} \int_0^\infty s^{n+1} \phi(s) \, ds \int_S (v \cdot u)^2 \lambda(u)^{-n-2} \, du \\ &= \hat{\sigma}^2 v \cdot [\Gamma_2 C]v. \end{aligned}$$

■

Note that if f is elliptically contoured, then the ellipsoid $\Gamma_2 f$ is, up to a scale factor, equal to the contour body C itself.

Definition: A probability distribution having density $f(x)$ on $\mathbf{R}^n$, has an $n \times n$ *Fisher information matrix* F associated with the family of distributions obtained by translating $f(x)$. The components of F are given by

$$F_{ij} = \int_{\mathbf{R}^n} \frac{\partial \log f}{\partial x_i}(x) \frac{\partial \log f}{\partial x_j}(x) f(x) \, dx \tag{21}$$

and the corresponding *Fisher information ellipsoid* by

$$\Gamma_{-2} f = E_{F^{-1}}. \tag{22}$$

Observe that Fisher information is well–defined only if the gradient of the square root of density function f is L^2.

Lemma 5: The Fisher information ellipsoid $\Gamma_{-2} f$ of a contoured probability distribution with probability density function f is given by

$$\Gamma_{-2} f = \tilde{\sigma} \Gamma_{-2} C,$$

where

$$\tilde{\sigma}^{-2} = \frac{1}{n}\int_{\mathbf{R}^n} \frac{(\phi'(\|x\|))^2}{\phi(\|x\|)}\,dx. \tag{23}$$

Proof: Given any $v \in \mathbf{R}^n$,

$$\begin{aligned}
& v \cdot [\Gamma_{-2} f]^{-1} v \\
= & \int_{\mathbf{R}^n} (v \cdot \nabla f(x))^2 (f(x))^{-1}\,dx \\
= & \frac{v_n}{V(C)} \int_S \int_0^\infty \frac{\phi'(r\lambda(u))^2}{\phi(r\lambda(u))} (v \cdot \nabla\lambda(u))^2 r^{n-1}\,dr\,du \\
= & \left(v_n \int_0^\infty \frac{\phi'(s)^2}{\phi(s)} s^{n-1}\,ds \right) \\
& \cdot \frac{1}{V(C)} \int_S (v \cdot \nabla\lambda(u))^2 \lambda(u)^{-n}\,du \\
= & \left(v_n \int_0^\infty \frac{\phi'(s)^2}{\phi(s)} s^{n-1}\,ds \right) \\
& \cdot \frac{n}{V(C)} \int_S \int_0^{(\lambda(u))^{-1}} (v \cdot \nabla\lambda(ru))^2 r^{n-1}\,dr\,du \\
= & \tilde{\sigma}^{-2} \frac{n}{V(C)} \int_C (v \cdot \nabla\lambda(x))^2\,dx \\
= & \tilde{\sigma}^{-2} v \cdot [\Gamma_{-2} C]^{-1} v
\end{aligned}$$

where we have used (10) and (16). ■

Again, observe that if f is elliptically contoured, then the ellipsoids $\Gamma_{-2}f$, $\Gamma_2 f$, and C are, up to scale factors, the same.

IV. Extremal contoured distributions

A. Maximum entropy contoured distributions

It is well–known that among all probability distributions with a given covariance matrix, the unique distribution with maximum entropy is the Gaussian. In particular,

Theorem 6: (Theorem 9.6.5 of [4], using our notation in Equations (19) and (13).) Given a probability density function f on $\mathbf{R}^n$,

$$h[f] \le \log\left[(2\pi e)^{\frac{n}{2}} \frac{V(\Gamma_2 f)}{v_n}\right],$$

with equality holding if and only if f is a Gaussian distribution function.

This theorem combined with Lemma 3 implies the following.

Proposition 7: The entropy of a contoured probability distribution with density function f given by (10) satisfies the inequality

$$h[f] \leq \log\left[(2\pi e\hat{\sigma}^2)^{\frac{n}{2}}\frac{V(C)}{v_n}\right],$$

where $\hat{\sigma}^2$ is given by (20) and equality holds if and only if the radial profile of f is

$$\phi(r) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}}e^{-r^2/2\hat{\sigma}^2}.$$

In other words, among all contoured distributions with *given* contour body C and radial variance, the entropy is maximal if and only if the radial profile is Gaussian.

B. Minimum entropy contoured distributions

It is well–known that the isoperimetric inequality for bodies in $\mathbf{R}^n$ can be proved by differentiating the Brunn–Minkowski inequality. Stam [11] and Dembo [12] showed that a similar argument applied to Shannon's entropy power inequality results in the following sharp lower bound for entropy in terms of Fisher information.

Theorem 8: Given a probability distribution with density function f on $\mathbf{R}^n$,

$$h[f] \geq \log\left[(2\pi e)^{\frac{n}{2}}\frac{V(\Gamma_{-2}f)}{v_n}\right], \tag{24}$$

with equality holding if and only if f is a Gaussian distribution.
Again, note that we have used (22) and (13) to state the theorem in terms of the quantities of this paper.

This inequality has many different equivalent forms and has been proven many times in its different guises. The 1–dimensional case is due to Stam [11] in 1959. The higher dimensional case was first proved by Weissler [13] in 1978. On the other hand, Beckner–Pearson [14] prove that (24) is equivalent to the logarithmic Sobolev inequality proved by Gross [15] in 1975. The logarithmic Sobolev inequality continues to be studied actively [16], [17], [18]. The equality condition in Theorem 8 was established by Carlen [19].

This theorem combined with Lemma 5 implies the following.

Proposition 9: The entropy of a contoured probability distribution with density function f given by (10) satisfies the inequality

$$h[f] \geq \log\left[(2\pi e\tilde{\sigma}^2)^{\frac{n}{2}}\frac{V(C)}{v_n}\right],$$

where $\tilde{\sigma}$ is given by (23). Equality holds if and only if the radial profile of f is

$$\phi(r) = \frac{1}{(2\pi\tilde{\sigma}^2)^{n/2}} e^{-r^2/2\tilde{\sigma}^2}.$$

In other words, among all contoured distributions with *given* contour and radial Fisher information, the entropy is minimal if and only if the radial profile is Gaussian.

V. Geometric inequalities and applications to Renyi entropy

A. A volume bound for the Legendre ellipsoid

Theorem 6 immediately implies the following well–known elementary geometric inequality.

Proposition 10: Given any contour body $C \subset \mathbf{R}^n$,

$$V(C) \leq V(\Gamma_2 C),$$

with equality holding if and only if C is an ellipsoid centered at the origin.

Proof: Let

$$\phi(r) = \frac{1}{(2\pi)^{n/2}} e^{-r^2/2},$$

f be given by (10), and f_ϕ by (9). By Lemma 3, Theorem 6, and Lemma 4,

$$\begin{aligned} \log V(C) &= \log v_n + h[f] - h[f_\phi] \\ &\leq \log V(\Gamma_2 f) \\ &= \log V(\Gamma_2 C) \end{aligned}$$

Moreover, equality holds if and only if f is a Gaussian with mean zero, which implies that C is an ellipsoid centered at the origin. ■

This result leads to the following characterization of contoured distributions with maximum Renyi entropy.

Corollary 11: Given $q \geq 0$, the q–Renyi entropy of a contoured probability distribution with density function f given by (10) satisfies the inequality

$$h_q[f] \leq h_q[f_\phi] + \log \frac{V(\Gamma_2 f)}{\hat{\sigma}^n v_n},$$

with equality holding if and only if C is an ellipsoid, where f_ϕ is given by (9) and $\hat{\sigma}$ is given by (20).

In other words, among all contoured distributions with a given radial profile and covariance matrix, the q–Renyi entropy is maximal if and only if the distribution is elliptically contoured.

B. A volume bound for $\Gamma_{-2}C$

Theorem 8 can be used to prove the following inequality that was recently established by Lutwak, Yang, and Zhang [10].

Proposition 12: Given any contour body $C \subset \mathbf{R}^n$,

$$V(\Gamma_{-2}C) \leq V(C),$$

with equality holding if and only if C is an ellipsoid centered at the origin.

Proof: Let ϕ, f, and f_ϕ be as in the proof of Proposition 10. By Lemma 3, Theorem 8, and Lemma 5,

$$\begin{aligned} \log V(C) &= \log v_n + h[f] - h[f_\phi] \\ &\geq \log V(\Gamma_{-2}f) \\ &= \log V(\Gamma_{-2}C). \end{aligned}$$

By the equality condition of Theorem 8, equality holds if and only if f is Gaussian with mean zero and therefore C is an ellipsoid centered at the origin. ■

This result leads to the following characterization of contoured distributions with given Fisher information and radial profile and minimal Renyi entropy.

Corollary 13: Given $q \geq 0$, the q–Renyi entropy of a contoured probability distribution with density function f given by (10) satisfies the inequality

$$h_q[f] \geq h_q[f_\phi] + \log \frac{V(\Gamma_{-2}f)}{\tilde{\sigma}^n v_n},$$

with equality holding if and only if C is an ellipsoid, where f_ϕ is given by (9) and $\tilde{\sigma}$ is given by (23).

In other words, among all contoured distributions with a given radial profile and Fisher information matrix, the q–Renyi entropy is minimal if and only if the distribution is elliptically contoured.

VI. Reverse information theoretic inequalities

Lutwak, Yang, and Zhang [10] show that if a contour body C is assumed to be origin–symmetric[2] and convex, the following reverse inequality holds:

Theorem 14: Given any origin–symmetric convex contour body $C \subset \mathbf{R}^n$,

$$V(\Gamma_{-2}C) \geq 2^{-n} v_n V(C),$$

with equality holding if and only if C is a parallelotope centered at the origin.

We call a probability density function $f(x)$ *origin–symmetric*, if for every $x \in \mathbf{R}^n$,

$$f(-x) = f(x).$$

Note that a contoured distribution is origin–symmetric if and only if its contour body is origin–symmetric and that an origin–symmetric contoured distribution is always balanced.

We will call a contoured distribution *convex*, if its contour body is convex.

Theorem 14, along with Lemma 3 and Lemma 5, immediately implies the following remarkable reverse inequality:

Theorem 15: Let f be the density function of an origin–symmetric convex contoured probability distribution, f_ϕ be given by (9), and $\tilde{\sigma}$ be given by (23). Then given any $q \geq 0$,

$$h_q[f] \leq h_q[f_\phi] + \log \frac{V(\Gamma_{-2}f)}{\tilde{\sigma}^n v_n} + \log \frac{2^n}{v_n} \ ,$$

with equality holding if and only if the contour body of f is a parallelotope.

This theorem, along with Theorem 8, shows that it is possible to bound the entropy of an origin–symmetric convex contoured distribution from below *and above* using the distribution's Fisher information. An interesting question is whether there is a larger class of probability distributions that can be called convex in some sense and that also satisfy Theorem 15. It is worth noting, however, that the class of origin–symmetric convex contoured distributions already contains most, if not all, commonly used explicit examples of multidimensional probability distributions.

Given Theorem 14, it is natural to ask what is the best possible constant c such that any origin–symmetric convex contour body $C \subset \mathbf{R}^n$ satisfies

$$V(\Gamma_2 C) \leq c^n V(C).$$

[2] A set $\Sigma \subset \mathbf{R}^n$ is *origin–symmetric*, if for any $x \in \Sigma$, $-x \in \Sigma$.

Neither this nor the equality condition is understood. In particular, an important unsolved conjecture in convex geometry is that c is a bounded function of the dimension n (see [20] for details). This conjecture has the following information theoretic formulation.

Conjecture 1: For each $q \geq 0$ there exists a constant c (independent of the dimension n) such that the following holds: Any density function f of an origin–symmetric convex contoured probability distribution on $\mathbf{R}^n$ satisfies

$$h_q[f] \geq h_q[f_\phi] + \log \frac{V(\Gamma_2 f)}{\hat{\sigma}^n v_n} - n \log c,$$

where f_ϕ is given by (9).

VII. Conclusion

We have demonstrated how basic information theoretic invariants of a contoured probability distribution are closely related to geometric invariants of its contour body and how known information theoretic inequalities imply known geometric inequalities. On the other hand, a recently established reverse geometric inequality yields a new reverse information theoretic inequality for origin–symmetric convex contoured distributions. An outstanding conjecture in convex geometry implies another new reverse inequality for origin–symmetric convex contoured distributions.

Given the close relationship between the information theoretic properties of contoured distributions and geometric properties of the associated contour bodies, we believe that further study of contoured distributions would have significant impact in both information theory and geometry.

We also believe that the observations and results established in this paper should extend in some form to more general probability distributions. In particular, there should be a general notion of a convex probability distribution for which the reverse inequalities described here hold.

Appendix

Three different normalized representations of a contoured distribution are presented below. The proofs are straightforward and therefore omitted.

The normalization given by Lemma 16 below is perhaps the most natural. It normalizes the contour body uniquely, so that information about the variance or the spread of the distribution is completely contained in the radial profile. Although this is perhaps a reasonable approach, it may be preferable to have the geometry of the contour body reflect the spread of the distribution. The decompositions described in Lemmas 17 and 18 normalize the radial profile uniquely, allowing size of the contour body to reflect the spread of the distribution. However this is at the expense of requiring additional assumptions on f. Lemma 19 takes an approach based on robust statistics. It normalizes the percentile associated with a contour body. The size of the contour body therefore does reflect the spread of the distribution, while no assumptions on f are required. It does, however, require fixing the value of a constant used in the normalization.

Normalizing the volume of the contour body C leads to the following result on the contoured distribution:

Lemma 16: Given a contoured density function f, there is a unique decomposition of the form

$$f(x) = \phi(\lambda_C(x)).$$

With this normalization, the volume of the contour body is equal to the volume of a unit ball.

Another possibility is to normalize the variance of the associated spherically contoured density function f_ϕ:

Lemma 17: Given a contoured density function f such that

$$\int_{\mathbf{R}^n} \|x\|^2 f(x)\, dx < \infty, \tag{25}$$

there exists a unique decomposition

$$f(x) = \frac{v_n}{V(C)} \phi(\lambda_C(x)),$$

where v_n is given by (1), such that the associated spherically contoured density function f_ϕ, as given by (9) satisfies

$$\int_{\mathbf{R}^n} \|x\|^2 f_\phi(x)\, dx = n.$$

We show in §III-A that the mean of a contoured distribution is proportional to the center of mass of the contour body, viewed as a body with constant density. The two vectors are equal, if the following normalization is used:

Lemma 18: Given a contoured density function f such that

$$\int_{\mathbf{R}^n} \|x\| f(x)\, dx < \infty, \tag{26}$$

there exists a unique decomposition

$$f(x) = \frac{v_n}{V(C)} \phi(\lambda_C(x)),$$

where v_n is given by (1), such that the spherically contoured density function f_ϕ as given by (9) satisfies

$$\int_{\mathbf{R}^n} \|x\| f_\phi(x)\, dx = 1.$$

If, however, the density f has fat tails, neither of the two lemmas above may apply. We can instead normalize the contour body using a given percentile:

Lemma 19: Let $0 < p < 1$. For each contoured distribution with density function f, there exists a unique decomposition

$$f(x) = \frac{v_n}{V(C)} \phi(\lambda_C(x)),$$

where v_n is given by (1), such that

$$\int_C f(x)\, dx \quad = \quad p.$$

References

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

[2] Max H. M. Costa and Thomas M. Cover, "On the similarity of the entropy power inequality and the Brunn–Minkowski inequality," *IEEE Trans. Inform. Theory*, vol. 30, no. 6, pp. 837–839, 1984.

[3] Thomas M. Cover, Amir Dembo, and Joy A. Thomas, "Information theoretic inequalities," *IEEE Transactions on Information Theory*, vol. 37, pp. 1501–1518, 1991.

[4] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wily–Interscience, 1991.

[5] William Beckner, "Inequalities in Fourier analysis," *Ann. of Math. (2)*, vol. 102, no. 1, pp. 159–182, 1975.

[6] Herm Jan Brascamp and Elliott H. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," *Advances in Math.*, vol. 20, no. 2, pp. 151–173, 1976.

[7] Kai-Tai Fang and Yao-Ting Zhang, *Generalized Multivariate Analysis*, Springer–Verlag, 1990.

[8] Yoshio Yamada, Saburo Tazaki, and Robert M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory*, vol. 26, no. 1, pp. 6–14, 1980.

[9] J. Lindenstrauss and V. Milman, *Local theory of normed spaces and convexity*, pp. 1149–1120, Birkhäuser Verlag, Basel, 1995, Papers from the Israel Seminar (GAFA) held 1992–1994.

[10] Erwin Lutwak, Deane Yang, and Gaoyong Zhang, "A new ellipsoid associated with convex bodies," *Duke Math. J.*, vol. 104, no. 3, pp. 375–390, 2000.

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

[12] Amir Dembo, "Information inequalities and uncertainty principles," Tech. Rep., Department of Statistics, Stanford University, 1991.

[13] Fred B. Weissler, "Logarithmic Sobolev inequalities for the heat–diffusion semigroup," *Transactions of the American Mathematical Society*, vol. 237, pp. 255–269, 1978.

[14] William Beckner and Michael Pearson, "On sharp Sobolev embedding and the logarithmic Sobolev inequality," *Bulletin of the London Mathematical Society*, vol. 30, pp. 80–84, 1998.

[15] L. Gross, "Logarithmic Sobolev inequalities," *American Journal of Mathematics*, vol. 97, pp. 1061–1083, 1975.

[16] William Beckner, "Pitt's inequality and the uncertainty principle," *Proceedings of the American Mathematical Society*, vol. 123, pp. 1897–1905, 1995.

[17] William Beckner, "Geometric proof of Nash's inequality," *International Mathematics Research Notices*, , no. 2, pp. 67–71, 1998.

[18] William Beckner, "Geometric asymptotics and the logarithmic Sobolev inequality," *Forum Mathematicum*, vol. 11, pp. 105–137, 1999.

[19] Eric A. Carlen, "Superadditivity of Fisher's information and logarithmic Sobolev inequalities," *Journal of Functional Analysis*, vol. 101, pp. 194–211, 1991.

[20] V. D. Milman and A. Pajor, "Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space," in *Geometric aspects of functional analysis (1987–88)*, pp. 64–104. Springer, Berlin, 1989.

Biographies of Authors

Onur G. Guleryuz was born in Ankara, Turkey in 1969. He received the BS degree in electrical engineering and the BS degree in physics from Bogazici University, Istanbul, Turkey in 1991, the MS degree in engineering and applied science from Yale University, New Haven, CT in 1992, and the PhD degree in electrical engineering from University of Illinois at Urbana-Champaign in 1997.

From August 1992 to August 1997, he was a research assistant in the Image Formation and Processing Group at the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. From 1997 to 2000, he was an Assistant Professor with the Department of Electrical Engineering at Polytechnic University, Brooklyn, NY. He is now with Epson Palo Alto Laboratory, Palo Alto, CA.

His research interests include image and video processing, multiresolutional analysis and representations, and source coding.

Erwin Lutwak received a BS, MS, and Ph.D. degree (all in Mathematics) from the Polytechnic University, where he is now Professor of Mathematics.

Deane Yang received a B.A. in mathematics and physics from University of Pennsylvania and Ph.D. in mathematics from Harvard University.

He has been an NSF Postdoctoral Fellow at the Courant Institute, a visiting faculty member of Columbia University and Université de Paris–Sud at Orsay. He has been a visiting member of l'Institut des Hautes Etudes Scientifiques on numerous occasions. He was an associate professor at Rice University and is now a full professor at Polytechnic University.

Gaoyong Zhang received the BS degree in mathematics from Wuhan University of Technology, the MS degree in mathematics from Wuhan University, Wuhan, China, and the Ph.D. degree in mathematics from Temple University, Philadelphia.

He has been a Rademacher Lecturer in the University of Pennsylvania, a member of the Institute for Advanced Study at Princeton, and a member of the Mathematical Sciences Research Institute at Berkeley. He has been an assistant professor and is now a full professor at Polytechnic University.