Affine moments of a random vector

Erwin Lutwak, Songjun Lv, Deane Yang, and Gaoyong Zhang

***Abstract*—An affine invariant p-th moment measure is defined for a random vector and used to prove sharp moment-entropy inequalities that are more general and stronger than standard moment-entropy inequalities.**

***Index Terms*—Moment, affine moment, information measure, information theory, entropy, Rényi entropy, Shannon entropy, Shannon theory, entropy inequality, Gaussian, generalized Gaussian.**

I. Introduction

The moment-entropy inequality states that a continuous random variable with given second moment has maximal Shannon entropy if and only if it is a Gaussian (see, for example, [15], Theorem 9.6.5). The Fisher information inequality states that a continuous random variable with given Fisher information has minimal Shannon entropy if and only if it is a Gaussian (see [40]). In [32] these classical inequalities were extended to sharp inequalities involving the Rényi entropy, p-th moment, and generalized Fisher information of a random variable, where the extremal distributions are no longer Gaussians but are fat-tailed distributions that the authors call *generalized Gaussians.*

There are different ways to extend the definition of a p-th moment to random vectors in $\mathbf{R}^n$. The most obvious one is to define it using the Euclidean norm of the random vector. This, however, assumes that the standard inner product on $\mathbf{R}^n$ gives the right invariant scalar measure of error or noise in the random vector. It is more appropriate to seek a definition of moment that does not rely on the standard inner product.

For example, if the moment of a random vector is defined using the standard Euclidean norm, then the extremals for the corresponding classical moment-entropy inequality are Gaussians with covariance matrix equal to a multiple of the identity. It is more desirable to define an invariant moment measure, where *any* Gaussian (or generalized Gaussian) is an extremal distribution and not just the one whose covariance matrix is a multiple of the identity.

One approach for defining an invariant moment measure, taken in [35], leads to the definition of the p-th moment matrix and the corresponding sharp moment-entropy inequalities.

Here, we introduce a different approach, where a scalar affine moment measure is defined by averaging 1-dimensional p-th moments obtained by projecting an n-dimensional random vector along each possible direction in $\mathbf{R}^n$ with respect to a given probability measure, and then optimizing this average over all probability measures with given p-th moment. An explicit integral formula for this moment is derived. We then establish sharp information inequalities giving a sharp lower bound of Rényi entropy in terms of the affine moment. These new affine moment-entropy inequalities imply the moment-entropy inequalities obtained in [32], [35].

In [26] an approach similar to the one taken here is used to define an affine invariant version of Fisher information, and a corresponding sharp Fisher information inequality is established.

It is worth noting that the affine moment measure introduced here, as well as the affine Fisher information studied in [26], is closely related to analogous affine invariant generalizations of the surface area of a convex body (see, for example, [9], [27], [30]). In fact, the results presented here are part of an ongoing effort by the authors to explore connections between information theory and geometry.

Previous results on this include a connection between the entropy power inequality in information theory and the Brunn-Minkowski inequality in convex geometry first demonstrated by Lieb [23] and also discussed by Costa and Cover [12]. This was developed further by Cover, Dembo, and Thomas [14] (also, see [15]). Also, see [10] for generalizations of various entropy and Fisher information inequalities related to mass transportation, and [4]–[8], [37], [38] for a new connection between affine information inequalities and log-concavity.

In view of this, the authors of [19] began to systematically explore connections between information theory and convex geometry. The goals are to both establish information-theoretic inequalities that are the counterparts of geometric inequalities and investigate possible applications of ideas in information theory to convex geometry. This has led to papers in both information theory (see [19], [26], [31], [32], [35]) and convex geometry (see [28], [29]). In the next section we follow the suggestion of a referee and provide a brief survey of this ongoing effort.

II. Information theoretic and geometric inequalities

The usual way of associating a random vector X in $\mathbf{R}^n$ with a compact convex set K in $\mathbf{R}^n$ is to define X as the uniform random vector in K. In [19], a different construction of random vectors associated to a star-shaped set was introduced. It was also shown that the information theoretic invariants of the distributions constructed, called *contoured distributions* are equivalent to geometric invariants of the convex set K. This provides a direct link between sharp information theoretic inequalities satisfied by contoured distributions and sharp geometric inequalities satisfied by convex sets.

Let K be a bounded star-shaped set in $\mathbf{R}^n$ about the origin. Its gauge function, $g_K : \mathbf{R}^n \to [0, \infty)$, is defined by

$$g_K(x) = \inf\{t > 0 : x \in tK\}.$$

E. Lutwak (elutwak@poly.edu), D. Yang (dyang@poly.edu), and G. Zhang (gzhang@poly.edu) are with the Department of Mathematics, Polytechnic Institute of New York University, and were supported in part by NSF Grant DMS-0706859. S. Lv (lvsongjun@126.com) is with the College of Mathematics and Computer Science, Chongqing Normal University, and was supported in part by Chinese NSF Grant 10801140.

The gauge function is homogeneous of degree 1.

A random vector X has a contoured distribution if its probability density function is given by

$$f_X(x) = \psi(g_K(x - x_0)),$$

where K is a bounded star-shaped set with respect to the point x_0, and ψ is a 1-dimensional function that we call the *radial profile*. If ψ is monotone, then the level sets of f_X are dilates of K with respect to x_0.

A straightforward calculation shows that the entropy $h(X)$ of X is given by

$$h(X) = c_0(\psi, n)V(K),$$

where $c_0(\psi, n)$ is a constant depending only on the radial profile ψ and the dimension n and $V(K)$ is the n-dimensional volume (i.e., n-dimensional Lebesgue measure) of K. Another calculation shows that the Fisher information $J(X)$ of X is given by

$$J(X) = c_1(\psi, n)S_2(K),$$

where $c_1(\psi, n)$ is another constant depending on ψ and n and

$$S_2(K) = \int_{\partial K} \frac{dS(x)}{x \cdot \nu(x)}$$

is called the L_2 surface area of K, where the outer unit normal vector $\nu(x)$ exists for almost every $x \in \partial K$ with respect to the $(n-1)$-dimensional Hausdorff measure dS.

The L_2 surface area $S_2(K)$ is a geometric invariant in the L_p Brunn-Minkowski theory in convex geometry. The usual surface area is viewed as the L_1 surface area in the L_p Brunn-Minkowski theory (see [25]). The formula of the Fisher information and the L_2 surface area implies that the classical information theory is closely associated to the L_2 Brunn-Minkowski theory. Let K be a convex body (compact convex set with nonempty interior) and X be a random vector in $\mathbf{R}^n$. Denote by B the n-dimensional unit ball and by Z the n-dimensional standard Gaussian random vector. The following variational formula,

$$\lim_{t \to 0^+} \frac{V(K +_2 \sqrt{t}B) - V(K)}{t} = \frac{1}{2}S_2(K),$$

where $+_2$ is the L_2 Minkowski addition (see [25]), and the de Bruijn's identity,

$$\lim_{t \to 0^+} \frac{h(X + \sqrt{t}Z) - h(X)}{t} = \frac{1}{2}J(X),$$

further illustrate the close connection of information theory and geometry.

These connections lead to the definition of a new ellipsoid in the L_2 Brunn-Minkowski theory which is the counterpart of the Fisher information matrix (see [28]), and thus a Cramer-Raó inequality for star-shaped sets was proved in [29]. The results were, in turn, applied to information theory to show new information-theoretic inequalities (see [19]).

It is then natural to investigate the counterpart of the L_p Brunn-Minkowski theory in information theory. The new ellipsoid discovered was shown in [33] to be the L_2 case of a family of ellipsoids, called L_p John ellipsoids, while the classical John ellipsoid (the ellipsoid of maximal volume inside a convex body) is the L_∞ case. This new result in convex geometry suggests that it is natural to define a concept of L_p Fisher information matrix as a corresponding object of the L_p John ellipsoid. The usual Fisher information matrix is the L_2 case. This was done in the recent paper [26]. An extension of the covariance matrix to L_p covariance matrix (also called p-moment matrix) was given earlier in the paper [35], which corresponds to another family of ellipsoids in geometry that contains the classical Legendre ellipsoid and is conceptually dual to the family of L_p John ellipsoids.

Affine isoperimetric inequalities are central in the L_p Brunn-Minkowski theory. The authors have been exploring their corresponding affine information-theoretic inequalities. See the survey papers [24] and [42] on affine isoperimetric inequalities in convex geometry.

For $p \geq 1$, $\lambda > \frac{n}{n+p}$, and independent random vectors X, Y in $\mathbf{R}^n$, the following moment-entropy inequality was proved in [31],

$$E(|X \cdot Y|^p) \geq c\, N_\lambda(X)^p N_\lambda(Y)^p, \tag{1}$$

where N_λ denotes the λ-Rényi entropy power, and c is the best constant that is attained when X, Y are certain generalized Gaussian random vectors. The affine isoperimetric inequality behind this moment-entropy inequality is an L_p extension of the well-known Blaschke-Santaló inequality in geometry (see [36]).

The Shannon entropy $h(X)$ and the λ-Rényi entropy power $N_\lambda(X)$ of a continuous random vector X in $\mathbf{R}^n$ are affine invariants, that is, they are invariant under linear transformations of random vectors. To establish affine information-theoretic inequalities as counterparts of affine isoperimetric inequalities, affine notions of Fisher information and moments as corresponding notions of affine surface areas are needed.

In [26], the authors introduced the notion of affine (p, λ)-Fisher information $\Psi_{p,\lambda}(X)$ of a random vector X in $\mathbf{R}^n$, which is an analogue of the L_p integral affine surface area of a convex body. It was shown that the following affine Fisher information and entropy inequality holds:

$$\Psi_{p,\lambda}(X) N_\lambda(X)^{p((\lambda-1)n+1)} \geq c, \tag{2}$$

where $1 \leq p < n$, $\lambda > 1 - \frac{1}{n}$, and c is the best constant that is attained when X is a generalized Gaussian random vector. This inequality is proved by using an L_p affine Sobolev inequality established in [30] (see also [41]). The L_p affine Sobolev inequality is stronger than the classical L_p Sobolev inequality and comes from the L_p Petty projection inequality established in [27] which is an important affine isoperimetric inequality in the L_p Brunn-Minkowski theory of convex geometry.

It is one of the purposes this paper to introduce the notion of *affine p-th moment* $M_p(X)$ of a random vector X in $\mathbf{R}^n$, and to establish an affine moment-entropy inequality. We shall prove the following theorem.

Theorem 1: If $1 \leq p < \infty$, $\lambda > \frac{n}{n+p}$, and X is a random vector in $\mathbf{R}^n$ with finite λ-Rényi entropy and p-th moment, then

$$M_p(X) \geq c\, N_\lambda(X)^p, \tag{3}$$

where c is the best constant that is attained only when X is a generalized Gaussian random vector.

The inequality (3) is proved by using (1). Thus, the affine isoperimetric inequality behind the affine moment-entropy inequality (3) is the L_p extension of the Blaschke-Santaló inequality.

III. Preliminaries

Let X be a random vector in $\mathbf{R}^n$ with probability density function f_X.

If A is an invertible $n \times n$ matrix, then

$$f_{AX}(y) = |A|^{-1} f_X(A^{-1}y), \tag{4}$$

where $|A|$ is the absolute value of the determinant of A.

A. The p-th moment of a random vector

For $p \in (0, \infty)$, the p-th moment of X is defined to be

$$E(|X|^p) = \int_{\mathbf{R}^n} |x|^p f_X(x) dx,$$

where $|\cdot|$ denotes the standard Euclidean norm and dx the standard Lebesgue measure in $\mathbf{R}^n$.

B. Entropy power

The Shannon entropy of X is given by

$$h(X) = -\int_{\mathbf{R}^n} f_X(x) \log f_X(x)\, dx.$$

For $\lambda > 0$, the λ-Rényi entropy power of X is defined to be

$$N_\lambda(X) = \begin{cases} \left(\displaystyle\int_{\mathbf{R}^n} f_X(x)^\lambda\, dx\right)^{\frac{1}{n(1-\lambda)}} & \text{if } \lambda \neq 1, \\ e^{\frac{1}{n}h(X)} & \text{if } \lambda = 1, \end{cases}$$

and the λ-Rényi entropy is

$$h_\lambda(X) = n \log N_\lambda(X).$$

By (4),

$$N_\lambda(AX) = |A|^{\frac{1}{n}} N_\lambda(X), \tag{5}$$

for any invertible matrix A.

IV. Generalized Gaussians

A. Definition

If $\alpha > 0$ and $\beta < \alpha/n$, the corresponding generalized standard Gaussian random vector $Z \in \mathbf{R}^n$ has density function

$$f_Z(x) = \begin{cases} a_{\alpha,\beta}\left(1 - \frac{\beta}{\alpha}|x|^\alpha\right)_+^{\frac{1}{\beta}-\frac{n}{\alpha}-1} & \text{if } \beta \neq 0, \\ a_{\alpha,0}e^{-|x|^\alpha/\alpha} & \text{if } \beta = 0, \end{cases}$$

where $t_+ = \max\{t, 0\}$ for $t \in \mathbf{R}$,

$$a_{\alpha,\beta} = \begin{cases} \dfrac{\frac{\alpha}{n}|\frac{\beta}{\alpha}|^{\frac{n}{\alpha}}\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}B(\frac{n}{\alpha}, 1-\frac{1}{\beta})} & \text{if } \beta < 0, \\ \dfrac{\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}\alpha^{\frac{n}{\alpha}}\Gamma(\frac{n}{\alpha}+1)} & \text{if } \beta = 0, \\ \dfrac{\frac{\alpha}{n}(\frac{\beta}{\alpha})^{\frac{n}{\alpha}}\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}B(\frac{n}{\alpha}, \frac{1}{\beta}-\frac{n}{\alpha})} & \text{if } \beta > 0, \end{cases}$$

$\Gamma(\cdot)$ denotes the gamma function, and $B(\cdot,\cdot)$ denotes the beta function. Any random vector that can be written as $W = AZ$, for an invertible matrix A, is called a *generalized Gaussian*. If $\alpha = 2$ and $\beta = 0$, then Z is the standard Gaussian random vector with mean 0 and variance matrix equal to the identity.

The functions in the generalized standard Gaussian distributions, which are also called Barenblatt functions, have been found to arise naturally as extremals for Sobolev type inequalities and sharp inequalities relating moment, Rényi entropy, and generalized Fisher information (see, for example, [1], [3], [10], [11], [13], [16], [17], [20], [26], [31], [32], [34], [35]). The usual Gaussians and Student distributions are generalized Gaussians. The whole class of generalized Gaussian distributions in this form were first studied in [31] as extremal distributions of moment-entropy inequalities. Many authors have studied special cases of generalized Gaussians (see, for example, [1]–[3], [13], [16], [18], [21], [22], [39]).

B. Information measures of generalized Gaussians

Given $0 < p < \infty$ and $\lambda > n/(n+p)$, set the parameters α and β of the standard generalized Gaussian Z so that

$$\alpha = p, \qquad \frac{1}{\beta} - \frac{n+\alpha}{\alpha} = \frac{1}{\lambda - 1}, \tag{6}$$

and $\beta = 0$ when $\lambda = 1$. We assume throughout this paper that p, λ, and the parameters α and β of the standard generalized Gaussian satisfy these equations.

The p-th moment of Z is given by

$$E(|Z|^p) = n, \tag{7}$$

and its Rényi entropy power by

$$N_\lambda(Z) = \begin{cases} a_{\alpha,\beta}^{-\frac{1}{n}}\left(1 - \dfrac{n\beta}{\alpha}\right)^{\frac{1}{n(1-\lambda)}} & \text{if } \lambda \neq 1, \\ a_{\alpha,0}^{-\frac{1}{n}}e^{\frac{1}{\alpha}} & \text{if } \lambda = 1. \end{cases} \tag{8}$$

See [35] for similar formulas. The following are sketches of calculation.

For $\beta < 0$ ($\frac{n}{n+\alpha} < \lambda < 1$), by polar coordinates and change of variable $(1 - \frac{\beta}{\alpha}r^\alpha) = \frac{1}{t}$, we have

$$\begin{aligned} &E(|Z|^\alpha) \\ &= a_{\alpha,\beta}\int_{\mathbb{R}^n}|x|^\alpha\left(1 - \frac{\beta}{\alpha}|x|^\alpha\right)_+^{\frac{1}{\beta}-\frac{n}{\alpha}-1} dx \\ &= a_{\alpha,\beta} n\omega_n \int_0^\infty \left(1 - \frac{\beta}{\alpha}r^\alpha\right)^{\frac{1}{\beta}-\frac{n}{\alpha}-1} r^{n+\alpha-1}dr \\ &= a_{\alpha,\beta} n\omega_n \Big(\frac{\alpha}{-\beta}\Big)^{\frac{n}{\alpha}+1}\frac{1}{\alpha}\int_0^1 t^{-\frac{1}{\beta}-1}(1-t)^{\frac{n}{\alpha}}dt \\ &= a_{\alpha,\beta}\frac{n\pi^{\frac{n}{2}}}{\alpha\Gamma(1+\frac{n}{2})}\Big(\frac{\alpha}{-\beta}\Big)^{\frac{n}{\alpha}+1}B\Big(\frac{1}{-\beta}, \frac{n}{\alpha}+1\Big) \\ &= n, \end{aligned}$$

and

$$
\begin{aligned}
&N_\lambda(Z)\\
&=\left(a_{\alpha,\beta}^{\lambda}\int_{\mathbb{R}^n}\left(1-\frac{\beta}{\alpha}|x|^\alpha\right)_+^{\lambda(\frac{1}{\beta}-\frac{n}{\alpha}-1)}dx\right)^{\frac{1}{n(1-\lambda)}}\\
&=a'_{\alpha,\beta}\left[n\omega_n\int_0^\infty\left(1-\frac{\beta}{\alpha}r^\alpha\right)_+^{\frac{1}{\beta}-\frac{n}{\alpha}}r^{n-1}dr\right]^{\frac{1}{n(1-\lambda)}}\\
&=a'_{\alpha,\beta}\left[\frac{n\omega_n}{\alpha}\Big(\frac{\alpha}{-\beta}\Big)^{\frac{n}{\alpha}}\int_0^1 t^{-\frac{1}{\beta}-1}(1-t)^{\frac{n}{\alpha}-1}dt\right]^{\frac{1}{n(1-\lambda)}}\\
&=a'_{\alpha,\beta}\left[\frac{n\omega_n}{\alpha}\Big(\frac{\alpha}{-\beta}\Big)^{\frac{n}{\alpha}}\frac{1}{\alpha}B\Big(\frac{1}{-\beta},\frac{n}{\alpha}\Big)\right]^{\frac{1}{n(1-\lambda)}}\\
&=a_{\alpha,\beta}^{-\frac{1}{n}}\Big(1-\frac{\beta n}{\alpha}\Big)^{\frac{1}{n(1-\lambda)}},
\end{aligned}
$$

where $a'_{\alpha,\beta}=a_{\alpha,\beta}^{-\frac{1}{n}(1+\frac{1}{\lambda-1})}$.

For $0<\beta<\frac{\alpha}{n}$ ($\lambda>1$), by polar coordinates and change of variable $\frac{\beta}{\alpha}r^\alpha=t$, we have

$$
\begin{aligned}
&E(|Z|^\alpha)\\
&=a_{\alpha,\beta}\int_{\mathbb{R}^n}|x|^\alpha\left(1-\frac{\beta}{\alpha}|x|^\alpha\right)_+^{\frac{1}{\beta}-\frac{n}{\alpha}-1}dx\\
&=a_{\alpha,\beta}n\omega_n\int_0^\infty\left(1-\frac{\beta}{\alpha}r^\alpha\right)_+^{\frac{1}{\beta}-\frac{n}{\alpha}-1}r^{n+\alpha-1}dr\\
&=a_{\alpha,\beta}n\omega_n\Big(\frac{\alpha}{\beta}\Big)^{\frac{n}{\alpha}+1}\frac{1}{\alpha}\int_0^1 t^{\frac{n}{\alpha}}(1-t)^{\frac{1}{\beta}-\frac{n}{\alpha}-1}dt\\
&=\frac{a_{\alpha,\beta}n\pi^{\frac{n}{2}}}{\alpha\Gamma(1+\frac{n}{2})}\Big(\frac{\alpha}{\beta}\Big)^{\frac{n}{\alpha}+1}B\Big(\frac{1}{\beta}-\frac{n}{\alpha},\frac{n}{\alpha}+1\Big)\\
&=n,
\end{aligned}
$$

and

$$
\begin{aligned}
&N_\lambda(Z)\\
&=\left(a_{\alpha,\beta}^{\lambda}\int_{\mathbb{R}^n}\left(1-\frac{\beta}{\alpha}|x|^\alpha\right)_+^{\lambda(\frac{1}{\beta}-\frac{n}{\alpha}-1)}dx\right)^{\frac{1}{n(1-\lambda)}}\\
&=a'_{\alpha,\beta}\left[n\omega_n\int_0^\infty\left(1-\frac{\beta}{\alpha}r^\alpha\right)_+^{\frac{1}{\beta}-\frac{n}{\alpha}}r^{n-1}dr\right]^{\frac{1}{n(1-\lambda)}}\\
&=a'_{\alpha,\beta}\left[\frac{n\omega_n}{\alpha}\Big(\frac{\alpha}{\beta}\Big)^{\frac{n}{\alpha}}\int_0^1 t^{\frac{n}{\alpha}-1}(1-t)^{\frac{1}{\beta}-\frac{n}{\alpha}}dt\right]^{\frac{1}{n(1-\lambda)}}\\
&=a'_{\alpha,\beta}\left[\frac{n\omega_n}{\alpha}\Big(\frac{\alpha}{\beta}\Big)^{\frac{n}{\alpha}}B\Big(1+\frac{1}{\beta}-\frac{n}{\alpha},\frac{n}{\alpha}\Big)\right]^{\frac{1}{n(1-\lambda)}}\\
&=a_{\alpha,\beta}^{-\frac{1}{n}}\Big(1-\frac{\beta n}{\alpha}\Big)^{\frac{1}{n(1-\lambda)}}.
\end{aligned}
$$

For $\beta=0$ ($\lambda=1$), by polar coordinates and change of variable $\frac{1}{\alpha}r^\alpha=t$, we have

$$
\begin{aligned}
&E(|Z|^\alpha)\\
&=a_{\alpha,0}\int_{\mathbb{R}^n}|x|^\alpha e^{-\frac{1}{\alpha}|x|^\alpha}dx\\
&=a_{\alpha,0}n\omega_n\int_0^\infty e^{-\frac{1}{\alpha}r^\alpha}r^{n+\alpha-1}dr\\
&=a_{\alpha,0}n\omega_n\alpha^{\frac{n}{\alpha}}\int_0^\infty e^{-t}t^{\frac{n}{\alpha}}dt\\
&=a_{\alpha,0}\frac{n\pi^{\frac{n}{2}}}{\Gamma(1+\frac{n}{2})}\alpha^{\frac{n}{\alpha}}\Gamma\Big(\frac{n}{\alpha}+1\Big)\\
&=n,
\end{aligned}
$$

$$
\begin{aligned}
&h(Z)\\
&=-\int_{\mathbb{R}^n}f_Z(x)\log f_Z(x)\,dx\\
&=-\int_{\mathbb{R}^n}f_Z(x)\log\left(a_{\alpha,0}e^{-\frac{|x|^\alpha}{\alpha}}\right)dx\\
&=-\int_{\mathbb{R}^n}f_Z(x)\log a_{\alpha,0}+f_Z(x)\Big(-\frac{|x|^\alpha}{\alpha}\Big)dx\\
&=-\log a_{\alpha,0}+\frac{1}{\alpha}E(|Z|^\alpha)\\
&=-\log a_{\alpha,0}+\frac{n}{\alpha},
\end{aligned}
$$

and thus,

$$N_1(Z)=a_{\alpha,0}^{-\frac{1}{n}}e^{\frac{1}{\alpha}}.$$

V. Notions of an affine moment

If G is the standard Gaussian random vector in $\mathbf{R}$ with mean 0 and variance matrix equal to the identity, then the classical moment-entropy inequality (see, for example, [15]) states that for a random vector X in $\mathbf{R}$,

$$\frac{E(|X|^2)}{N_1(X)^2}\geq\frac{E(|G|^2)}{N_1(G)^2}, \tag{9}$$

with equality if and only if $X=tG$, for some $t\in\mathbf{R}\backslash\{0\}$. In [31], [32], [35], this was extended to the following inequality for the λ-Rényi entropy and p-th moment.

Theorem 2: If $p\in(0,\infty)$, $\lambda>n/(n+p)$, and $X\in\mathbf{R}^n$ is a random vector such that $N_\lambda(X),E(|X|^p)<\infty$, then

$$\frac{E(|X|^p)}{N_\lambda(X)^p}\geq\frac{E(|Z|^p)}{N_\lambda(Z)^p},$$

with equality if and only if there exists $t>0$ such that $X=tZ$, where Z is the standard generalized Gaussian with parameters α and β satisfying (6).

In [35], an affine p-th moment of a random vector X is defined by

$$m_p(X)=\inf\{E(|AX|^p)\ :\ A\in\mathrm{SL}(n)\},$$

and the following affine moment-entropy inequality was shown.

Theorem 3: If $p\in(0,\infty)$, $\lambda>n/(n+p)$, and X is a random vector in $\mathbf{R}^n$ satisfying $N_\lambda(X),E(|X|^p)<\infty$, then

$$\frac{m_p(X)}{N_\lambda(X)^p}\geq\frac{E(|Z|^p)}{N_\lambda(Z)^p},$$

with equality if and only if $X = TZ$ for the standard generalized Gaussian Z and some $T \in \mathrm{GL}(n)$.

Theorem 3 is formally stronger than Theorem 2, but the two are equivalent. Theorem 3 is an affine formulation of Theorem 2. The affine moment $m_p(X)$ has no explicit formula in terms of the density of X. This makes the calculation difficult. Is there a notion of affine moments that has an explicit formula and also gives an essentially stronger moment-entropy inequality than those in Theorems 2 and 3?

The definition of $m_p(X)$ can be formulated differently as follows: Let $\mathscr{F}$ be the class of norms on $\mathbf{R}^n$ that is given by

$$\mathscr{F} = \{\|\cdot\|_A : A \in \mathrm{SL}(n)\},$$

where $\|\cdot\|_A$ is defined by $\|x\|_A = |Ax|$, $x \in \mathbf{R}^n$. Then

$$m_p(X) = \inf_{\|\cdot\|_A \in \mathscr{F}} E(\|X\|_A^p).$$

We shall use a larger class of norms than $\mathscr{F}$ to define the affine moment $M_p(X)$. The norms are generated by the p-cosine transforms of density functions. We shall give an explicit formula for the affine moment $M_p(X)$ and establish the affine moment-entropy inequality in Theorem 1 which is essentially stronger than Theorems 2 and 3.

A similar approach was used in [26] to define a notion of affine Fisher information.

VI. The affine p-th moment of a random vector

A. Definition

A random vector X in $\mathbf{R}^n$ with density g is said to have finite p-moment for $p > 0$, if

$$\int_{\mathbf{R}^n} |x|^p g(x)\, dx < \infty.$$

The p-cosine transform of a random vector Y with finite p-th moment and density g defines the following norm on $\mathbf{R}^n$,

$$\|x\|_{Y,p} = \left(\int_{\mathbf{R}^n} |x \cdot y|^p g(y)\, dy \right)^{\frac{1}{p}}, \ x \in \mathbf{R}^n. \tag{10}$$

If $p > 0$ and X is a random vector, then we define the *affine p-th moment of* X to be

$$M_p(X) = \inf_{N_\lambda(Y)=c_1} E(\|X\|_{Y,p}^p), \tag{11}$$

where each random vector Y is independent of X and has finite p-th moment and λ-Rényi entropy power equal to a constant c_1 which will be chosen appropriately later.

The definition above appears to depend on the parameter λ, but by Theorem 5 and (14), which are stated in Sections VI-B and VI-D, the value of $M_p(X)$ is in fact independent of λ when the constant c_1 is properly chosen. We also show below that the infimum in the definition above is achieved, and the affine p-th moment $M_p(X)$ is invariant under volume-preserving linear transformations of the random vector X.

B. An integral representation for affine moments

The following is a special case of the dual Minkowski inequality for random vectors established in [31], Lemma 4.1.

Lemma 4: If $p > 0$, $\lambda > \frac{n}{n+p}$, and X and Y are independent random vectors in $\mathbf{R}^n$ with finite p-th moment, then

$$\int_{\mathbf{R}^n} E(|y \cdot X|^p) g(y)\, dy \geq N_\lambda(Y)^p \left(a_1 \int_{S^{n-1}} E(|u \cdot X|^p)^{-\frac{n}{p}}\, du \right)^{-\frac{p}{n}},$$

where g is the density of Y, S^{n-1} is the unit sphere in $\mathbf{R}^n$, du denotes the Lebesgue measure on S^{n-1},

$$a_1 = \begin{cases} \dfrac{a_0}{n} B\left(\dfrac{n}{p}, \dfrac{1}{1-\lambda} - \dfrac{n}{p}\right) & \text{if } \lambda < 1, \\ \dfrac{1}{n}\Big(\dfrac{pe}{n}\Big)^{\frac{n}{p}} \Gamma\left(1 + \dfrac{n}{p}\right) & \text{if } \lambda = 1, \\ \dfrac{a_0}{n} B\left(\dfrac{n}{p}, \dfrac{\lambda}{\lambda - 1}\right) & \text{if } \lambda > 1, \end{cases}$$

and

$$a_0 = \frac{n}{p} \left(1 + \frac{n(\lambda-1)}{p\lambda} \right)^{\frac{1}{\lambda-1}} \left| 1 + \frac{p\lambda}{n(\lambda-1)} \right|^{\frac{n}{p}}.$$

Equality is attained, if the density of Y is given by

$$g(y) = \begin{cases} b(1 + a\|y\|^p)^{\frac{1}{\lambda-1}} & \text{if } \lambda < 1, \\ be^{-a\|y\|^p} & \text{if } \lambda = 1, \\ b(1 - a\|y\|^p)_+^{\frac{1}{\lambda-1}} & \text{if } \lambda > 1, \end{cases} \tag{12}$$

for a, $b > 0$, where the norm $\|\cdot\|$ is given by $\|y\|^p = E(|y \cdot X|^p)$ for each $y \in \mathbf{R}^n$.

The following is the integral representation for the affine p-th moment.

Theorem 5: If $p > 0$ and X is a random vector with finite p-th moment and density f, then

$$M_p(X) = \left[c_0 \int_{S^{n-1}} E(|u \cdot X|^p)^{-\frac{n}{p}} du \right]^{-\frac{p}{n}},$$

where $c_0 = \frac{a_1}{c_1^{\frac{n}{p}}}$.

Proof: If Y is a random vector such that $N_\lambda(Y) = c_1$, then by the Fubini theorem and Lemma 4,

$$\begin{aligned} &E(\|X\|_{Y,p}^p) \\ &= \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} |x \cdot y|^p g(y)\, dy\, f(x)\, dx \\ &= \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} |x \cdot y|^p f(x)\, dx\, g(y)\, dy \\ &= \int_{\mathbf{R}^n} E(|y \cdot X|^p) g(y)\, dy \\ &\geq \left[c_0 \int_{S^{n-1}} E(|u \cdot X|^p)^{-\frac{n}{p}}\, du \right]^{-\frac{p}{n}}. \end{aligned}$$

Moreover, by the equality condition of Lemma 4, equality is attained if Y is a random vector whose density is given by

(12), normalized so that $N_\lambda(Y) = c_1$. Therefore,

$$\begin{aligned} M_p(X) &= \inf_{N_\lambda(Y)=c_1} E(\|X\|_{Y,p}^p) \\ &= \left[c_0 \int_{S^{n-1}} E(|u \cdot X|^p)^{-\frac{n}{p}}\, du\right]^{-\frac{p}{n}}. \end{aligned}$$

∎

C. Affine invariance of the affine p-th moment

Theorem 6: If X is a random vector in $\mathbf{R}^n$ with finite p-th moment, then

$$M_p(AX) = M_p(X),$$

for each $A \in \mathrm{SL}(n)$.

Proof: If $A \in \mathrm{GL}(n)$, then it follows by (10) that

$$\begin{aligned} &E(\|AX\|_{Y,p}^p) \\ &= \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} |Ax \cdot y|^p g(y)\, dy\, f(x)\, dx \\ &= \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} |x \cdot A^t y|^p g(y)\, dy\, f(x)\, dx \\ &= E(\|X\|_{A^tY}^p). \end{aligned}$$

Thus, if $A \in$ SL(n), then by (5),

$$\begin{aligned} M_p(AX) &= \inf_{N_\lambda(Y)=c_1} E(\|AX\|_{Y,p}^p) \\ &= \inf_{N_\lambda(Y)=c_1} E(\|X\|_{A^tY,p}^p) \\ &= \inf_{N_\lambda(A^tY)=c_1} E(\|X\|_{A^tY,p}^p) \\ &= M_p(X). \end{aligned}$$

∎

D. The affine p-th moment of a spherically contoured random vector

Denote the volume of the unit ball in $\mathbf{R}^n$ by

$$\omega_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(1 + \frac{n}{2})},$$

and observe that

$$\int_{S^{n-1}} du = n\omega_n. \tag{13}$$

Let

$$\omega_{n,p} = \int_{S^{n-1}} |u \cdot e_n|^p\, du = \frac{2\pi^{\frac{n-1}{2}} \Gamma(\frac{p+1}{2})}{\Gamma(\frac{n+p}{2})},$$

where e_n is a fixed unit vector, for example, $e_n = (0, \ldots, 0, 1)$.

We choose the constant c_1 so that the constant $c_0 = \frac{a_1}{c_1^n}$ is given by

$$c_0 = \frac{1}{n\omega_n} \left(\frac{\omega_{n,p}}{n\omega_n}\right)^{\frac{n}{p}} = \frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}} \left(\frac{\Gamma(\frac{n}{2})\Gamma(\frac{p+1}{2})}{\pi^{\frac{1}{2}}\Gamma(\frac{n+p}{2})}\right)^{\frac{n}{p}}. \tag{14}$$

A random vector X is called *spherically contoured* if its density f can be written as $f(x) = F(|x|)$, $x \in \mathbf{R}^n$, where $F : [0, \infty) \to [0, \infty)$.

Lemma 7: If X is a spherically contoured random vector with finite p-th norm and density given by $f(x) = F(|x|)$, $x \in \mathbf{R}^n$, then

$$M_p(X) = n\omega_n \int_0^\infty F(\rho)\rho^{p+n-1}\, d\rho. \tag{15}$$

Proof: For each $u \in S^{n-1}$,

$$\begin{aligned} E(|u \cdot X|^p) &= \int_{\mathbf{R}^n} |u \cdot x|^p\, F(|x|)\, dx \\ &= \int_0^\infty \int_{S^{n-1}} |u \cdot v|^p\, dv\, F(\rho)\rho^{p+n-1}\, d\rho \\ &= \omega_{n,p} \int_0^\infty F(\rho)\rho^{p+n-1}\, d\rho. \end{aligned}$$

By Theorem 5,

$$\begin{aligned} M_p(X) &= \left[c_0 \int_{S^{n-1}} E(|u \cdot X|^p)^{-\frac{n}{p}}\, du\right]^{-\frac{p}{n}} \\ &= n\omega_n \int_0^\infty F(\rho)\rho^{p+n-1}\, d\rho. \end{aligned}$$

∎

E. Affine versus Euclidean

Lemma 8: If $p > 0$ and X is a random vector in $\mathbf{R}^n$, then

$$M_p(X) \leq E(|X|^p). \tag{16}$$

Equality holds if and only if the function $v \mapsto E(|v \cdot X|^p)$ is constant for $v \in S^{n-1}$. In particular, equality holds if X is spherically contoured.

Proof: If f is the density of X and $u \in S^{n-1}$, then

$$\begin{aligned} &\int_{S^{n-1}} E(|u \cdot X|^p)\, du \\ &= \int_{S^{n-1}} \int_{\mathbf{R}^n} |u \cdot x|^p f(x)\, dx\, du \\ &= \int_{\mathbf{R}^n} \left(\int_{S^{n-1}} \left|u \cdot \frac{x}{|x|}\right|^p du\right) |x|^p f(x)\, dx \\ &= \omega_{n,p} E(|X|^p). \end{aligned}$$

Therefore, by Theorem 5, (13), and Hölder's inequality,

$$\begin{aligned} &(n\omega_n c_0)^{\frac{p}{n}} M_p(X) \\ &= \left[\frac{1}{n\omega_n} \int_{S^{n-1}} E(|u \cdot X|^p)^{-\frac{n}{p}}\, du\right]^{-\frac{p}{n}} \\ &\leq \frac{1}{n\omega_n} \int_{S^{n-1}} E(|u \cdot X|^p)\, du \\ &= \frac{\omega_{n,p}}{n\omega_n} E(|X|^p). \end{aligned}$$

The equality condition follows by the equality condition of Hölder's inequality. ∎

By the theorem above and (7), we get the following.

Corollary 9: If Z is the standard generalized Gaussian, then

$$M_p(Z) = n.$$

VII. Affine p-th moment-entropy inequalities

A. Proof of Theorem 1

The following bilinear moment-entropy inequality is established in [31].

Theorem 10: Let $p \geq 1$, $\lambda > n/(n+p)$. There exists a constant $c > 0$ such that if X and Y are independent random vectors in $\mathbf{R}^n$ with finite p-th moments, then

$$E(|X \cdot Y|^p) \geq c\, N_\lambda(X)^p N_\lambda(Y)^p,$$

with equality holding if and only if X and Y are certain generalized Gaussians.

We use this theorem to establish the following affine moment-entropy inequality, which is Theorem 1.

Theorem 11: If $p \in [1, \infty)$, $\lambda \in (n/(n+p), \infty)$, and X is a random vector in $\mathbf{R}^n$ with finite λ-Rényi entropy and p-th moment, then

$$\frac{M_p(X)}{N_\lambda(X)^p} \geq \frac{M_p(Z)}{N_\lambda(Z)^p}, \tag{17}$$

with equality if and only if X is a generalized Gaussian.

Proof: If Y is independent of X and has finite p-th moment and Rényi entropy power $N_\lambda(Y) = c_1$, then by (10) and Theorem 10,

$$\begin{aligned} E(\|X\|_{Y,p}^p) &= E(|X \cdot Y|^p) \\ &\geq c\, c_1^p N_\lambda(X)^p. \end{aligned}$$

The desired inequality (17) now follows by the definition (11) of $M_p(X)$. The equality condition follows by the equality conditions of Lemma 4 and Theorem 10 (or (7) and Corollary 9). ∎

B. Affine implies Euclidean

Proposition 12: The affine moment-entropy inequality in Theorem 11 is stronger than the Euclidean moment-entropy inequality in Theorem 2.

Proof: Observe that equality holds in Lemma 8 for a standard generalized Gaussian random vector Z because it is spherically contoured. By Lemma 8, Theorem 11, and the equality condition of Lemma 8,

$$\begin{aligned} \frac{E(|X|^p)}{N_\lambda(X)^p} &\geq \frac{M_p(X)}{N_\lambda(X)^p} \\ &\geq \frac{M_p(Z)}{N_\lambda(Z)^p} \\ &= \frac{E(|Z|^p)}{N_\lambda(Z)^p}. \end{aligned}$$

Therefore, the Euclidean moment-entropy inequality in Theorem 2 is weaker than the affine moment-entropy inequality in Theorem 11. ∎

Acknowledgment

The authors would like to thank the referees for their helpful comments.

References

[1] S. Amari, *Differential-geometrical methods in statistics*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, 1985, vol. 28.

[2] A. Andai, "On the geometry of generalized Gaussian distributions," *J. Multivariate Anal.*, vol. 100, no. 4, pp. 777–793, 2009.

[3] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 99–105, 1996.

[4] A. D. Barbour, O. Johnson, I. Kontoyiannis, and M. Madiman, "Compound Poisson approximation via information functionals," *Electron. J. Probab.*, vol. 15, pp. 1344–1368, 2010.

[5] S. Bobkov and M. Madiman, "Concentration of the information in data with log-concave distributions," *Ann. Probab.*, vol. 39, no. 4, pp. 1528–1543, 2011.

[6] S. Bobkov and M. Madiman, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 4940–4954, 2011.

[7] S. Bobkov and M. Madiman, "Dimensional behaviour of entropy and information," *C. R. Math. Acad. Sci. Paris*, vol. 349, no. 3-4, pp. 201–204, 2011.

[8] S. Bobkov and M. Madiman, "Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures," *J. Funct. Anal.*, vol. 262, no. 7, pp. 3309–3339, 2012.

[9] S. Campi and P. Gronchi, "The L^p-Busemann-Petty centroid inequality," *Adv. Math.*, vol. 167, pp. 128–141, 2002.

[10] D. Cordero-Erausquin, W. Gangbo, and C. Houdré, "Inequalities for generalized entropy and optimal transportation," *In Recent advances in the theory and applications of mass transport*, vol. 353, Contemp. Math., pp. 73–94. Amer. Math. Soc., Providence, RI, 2004.

[11] D. Cordero-Erausquin, B. Nazaret, and C. Villani, "A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities," *Adv. Math.*, vol. 182, pp. 307–332, 2004.

[12] M. H. M. Costa and T. M. Cover, "On the similarity of the entropy power inequality and the Brunn-Minkowski inequality," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 837–839, Nov. 1984.

[13] J. A. Costa, A. O. Hero, and C. Vignat, "A characterization of the multivariate distributions maximizing Renyi entropy," in *Proceedings of 2002 IEEE International Symposium on Information Theory*, 2002, p. 263.

[14] T. M. Cover, A. Dembo, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1501–1518, Nov. 1991.

[15] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley & Sons Inc., 1991, a Wiley-Interscience Publication.

[16] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[17] M. Del Pino and J. Dolbeault, "Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions," *J. Math. Pures Appl. (9)*, vol. 81, pp. 847–875, 2002.

[18] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, London, U.K.: Chapman & Hall, 1990, vol. 36, Monographs on Statistics and Applied Probability.

[19] O. G. Guleryuz, E. Lutwak, D. Yang, and G. Zhang, "Information-theoretic inequalities for contoured probability distributions," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2377–2383, Aug. 2002.

[20] O. Johnson and C. Vignat, "Some results concerning maximum Rényi entropy distributions," *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 43, no. 3, pp. 339–351, 2007.

[21] C. P. Kitsos and N. K. Tavoularis, "Logarithmic Sobolev inequalities for information measures," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2554–2561, 2009.

[22] C. Kitsos and N. Tavoularis, "Newentropy type information measures," in Proc. 31st Int. Conf. Inf. Technol. Interfaces, pp. 255–260, 2009.

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

[24] E. Lutwak, "Selected affine isoperimetric inequalities," *Handbook of Convex Geometry*, edited by Gruber and Wills, Elsevier, North- Holland, 1993, pp. 151–176.

[25] E. Lutwak, "The Brunn-Minkowski-Firey theory, I: Mixed volumes and the Minkowski problem," *J. Diff. Geom.*, vol. 38, pp. 131–150, 1993.

[26] E. Lutwak, S. Lv, D. Yang, and G. Zhang, "Extensions of Fisher Information and Stam's Inequality," *IEEE Trans. Inform. Theory*, vol. 58, pp. 1319–1327, 2012.

[27] E. Lutwak, D. Yang, and G. Zhang, "L_p affine isoperimetric inequalities," *J. Differential Geom.*, vol. 56, pp. 111–132, 2000.

[28] E. Lutwak, D. Yang, and G. Zhang, “A new ellipsoid associated with convex bodies,” *Duke Math. J.* vol. 104, pp. 375–390, 2000.
[29] E. Lutwak, D. Yang, and G. Zhang, “The Cramer-Rao inequality for star bodies,” *Duke Math. J.* vol. 112, pp. 59–81, 2002.
[30] E. Lutwak, D. Yang, and G. Zhang, “Sharp affine L_p Sobolev inequalities,” *J. Differential Geom.*, vol. 62, pp. 17–38, 2002.
[31] E. Lutwak, D. Yang, and G. Zhang, “Moment–entropy inequalities,” *Annals of Probability*, vol. 32, pp. 757–774, 2004.
[32] E. Lutwak, D. Yang, and G. Zhang, “Cramer-Raó and moment-entropy inequalities for Renyi entropy and generalized Fisher information,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 473–478, 2005.
[33] E. Lutwak, D. Yang, and G. Zhang, “L_p John ellipsoids,” Proc. London Math. Soc. vol. 90, pp. 497–520, 2005.
[34] E. Lutwak, D. Yang, and G. Zhang, “Optimal Sobolev norms and the L^p Minkowski problem,” *Int. Math. Res. Not.*, 2006, Art. ID 62987, 21 pp.
[35] E. Lutwak, D. Yang, and G. Zhang, “Moment-Entropy Inequalities for a Random Vector,” *IEEE Trans. Inform. Theory*, vol. 53, pp. 1603–1607, 2007.
[36] E. Lutwak and G. Zhang, “Blaschke-Santaló inequalities,” *J. Differential Geom.*, vol. 47, pp. 1–16, 1997.
[37] M. Madiman and A. Barron, “Generalized entropy power inequalities and monotonicity properties of information,” *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.
[38] M. Madiman and P. Tetali, “Information inequalities for joint distributions, with interpretations and applications,” *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2699–2713, 2010.
[39] S. Nadarajah, “TheKotz-type distribution with applications,” *Statistics*, vol. 37, no. 4, pp. 341–358, 2003.
[40] A. J. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Information and Control*, vol. 2, pp. 101–112, 1959.
[41] G. Zhang, “The affine Sobolev inequality,” *J. Differential Geom.*, vol. 53, pp. 183–202, 1999.
[42] G. Zhang, “New affine isoperimetric inequalities,” in: *ICCM 2007*, vol. II, pp. 239–267.

Erwin Lutwak is Professor of Mathematics at Polytechnic Institute of NYU. He received his B.S. and M.S. in Mathematics from this institution when it was named Polytechnic Institute of Brooklyn and his Ph.D. in Mathematics when it was named Polytechnic Institute of New York.

Songjun Lv is Associate Professor of Mathematics at Chongqing Normal University. He received his B.S. and M.S. in Mathematics from Hunan Normal University, and his Ph.D. in Mathematics from Shanghai University.

Deane Yang is Professor of Mathematics at Polytechnic Institute of New York University. He received his B.A. in Mathematics and Physics from University of Pennsylvania and Ph.D. in Mathematics from Harvard University.

Gaoyong Zhang is Professor of Mathematics at Polytechnic Institute of New York University. He received his B.S. in Mathematics from Wuhan University of Science and Technology, M.S. in Mathematics from Wuhan University, and Ph.D. in Mathematics from Temple University.