f-Divergence for convex bodies *

Elisabeth M. Werner †

Abstract

We introduce f-divergence, a concept from information theory and statistics, for convex bodies in $\mathbb{R}^n$. We prove that f-divergences are $SL(n)$ invariant valuations and we establish an affine isoperimetric inequality for these quantities. We show that generalized affine surface area and in particular the L_p affine surface area from the L_p Brunn Minkowski theory are special cases of f-divergences.

1 Introduction.

In information theory, probability theory and statistics, an f-divergence is a function $D_f(P, Q)$ that measures the difference between two probability distributions P and Q. The divergence is intuitively an average, weighted by the function f, of the odds ratio given by P and Q. These divergences were introduced independently by Csiszár [2], Morimoto [37] and Ali & Silvey [1]. Special cases of f-divergences are the Kullback Leibler divergence or relative entropy and the Rényi divergences (see Section 1).

Due to a number of highly influential works (see, e.g., [4] - [11], [14], [15], [19], [20], [22] - [27], [29], [31], [34] - [36], [38], [42], [43] - [54], [56] - [58]), the L_p-Brunn-Minkowski theory is now a central part of modern convex geometry. A fundamental notion within this theory is L_p affine surface area, introduced by Lutwak in the ground breaking paper [26].

It was shown in [52] that L_p affine surface areas are entropy powers of Rényi divergences of the cone measures of a convex body and its polar,

*Keywords: f-divergence, relative entropy, affine surface area. 2010 Mathematics Subject Classification: 52A20, 53A15

†Partially supported by an NSF grant, a FRG-NSF grant and a BSF grant

thus establishing further connections between information theory and convex geometric analysis. Further examples of such connections are e.g. several papers by Lutwak, Yang, and Zhang [28, 30, 32, 33] and the recent article [39] where it is shown how relative entropy appears in convex geometry.

In this paper we introduce f-divergences to the theory of convex bodies and thus strengthen the already existing ties between information theory and convex geometric analysis. We show that generalizations of the L_p affine surface areas, the L_ϕ and L_ψ affine surface areas introduced in [23] and [21], are in fact f-divergences for special functions f. We show that f-divergences are $SL(n)$ invariant valuations and establish an affine isoperimetric inequality for these quantities. Finally, we give geometric characterizations of f-divergences.

Usually, in the literature, f-divergences are considered for convex functions f. A similar theory with the obvious modifications can be developed for concave functions. Here, we restrict ourselves to consider the convex setting.

Further Notation.

We work in $\mathbb{R}^n$, which is equipped with a Euclidean structure $\langle \cdot, \cdot \rangle$. We write B_2^n for the Euclidean unit ball centered at 0 and S^{n-1} for the unit sphere. Volume is denoted by $| \cdot |$ or, if we want to emphasize the dimension, by $\mathrm{vol}_d(A)$ for a d-dimensional set A.

Let $\mathcal{K}_0$ be the space of convex bodies K in $\mathbb{R}^n$ that contain the origin in their interiors. Throughout the paper, we will only consider such K. For $K \in \mathcal{K}_0$, $K^\circ = \{y \in \mathbb{R}^n : \langle x, y \rangle \leq 1 \text{ for all } x \in K\}$ is the polar body of K. For a point $x \in \partial K$, the boundary of K, $N_K(x)$ is the outer unit normal in x to K and $\kappa_K(x)$, or, in short κ, is the (generalized) Gauss curvature in x. We write $K \in C^2_+$, if K has C^2 boundary ∂K with everywhere strictly positive Gaussian curvature κ_K. By μ or μ_K we denote the usual surface area measure on ∂K and by σ the usual surface area measure on S^{n-1}.

Let K be a convex body in $\mathbb{R}^n$ and let $u \in S^{n-1}$. Then $h_K(u)$ is the support function of K in direction $u \in S^{n-1}$, and $f_K(u)$ is the curvature function, i.e. the reciprocal of the Gaussian curvature $\kappa_K(x)$ at the point $x \in \partial K$ that has u as outer normal.

2 f-divergences.

Let (X, μ) be a measure space and let $dP = pd\mu$ and $dQ = qd\mu$ be probability measures on X that are absolutely continuous with respect to the measure μ. Let $f : (0, \infty) \to \mathbb{R}$ be a convex function. The $*$-adjoint function $f^* : (0, \infty) \to \mathbb{R}$ of f is defined by (e.g. [17])

$$f^*(t) = tf(1/t), \quad t \in (0, \infty). \tag{1}$$

It is obvious that $(f^*)^* = f$ and that f^* is again convex if f is convex. Csiszár [2], and independently Morimoto [37] and Ali & Silvery [1] introduced the f-divergence $D_f(P, Q)$ of the measures P and Q which, for a convex function $f : (0, \infty) \to \mathbb{R}$ can be defined as (see [17])

$$\begin{aligned} D_f(P, Q) &= \int_{\{pq>0\}} f\left(\frac{p}{q}\right) qd\mu + f(0)\ Q\left(\{x \in X : p(x) = 0\}\right) \\ &+ f^*(0)\ P\left(\{x \in X : q(x) = 0\}\right), \end{aligned} \tag{2}$$

where

$$f(0) = \lim_{t \downarrow 0} f(t) \quad \text{and} \quad f^*(0) = \lim_{t \downarrow 0} f^*(t). \tag{3}$$

We make the convention that $0 \cdot \infty = 0$.

Please note that

$$D_f(P, Q) = D_{f^*}(Q, P). \tag{4}$$

With (3) and as

$$f^*(0)\ P\left(\{x \in X : q(x) = 0\}\right) = \int_{\{q=0\}} f^*\left(\frac{q}{p}\right) pd\mu = \int_{\{q=0\}} f\left(\frac{p}{q}\right) qd\mu,$$

we can write in short

$$D_f(P, Q) = \int_X f\left(\frac{p}{q}\right) qd\mu. \tag{5}$$

For particular choices of f we get many common divergences. E.g. for $f(t) = t \ln t$ with $*$-adjoint function $f^*(t) = -\ln t$, the f-divergence is the classical *information divergence*, also called *Kullback-Leibler divergence* or *relative entropy* from P to Q (see [3])

$$D_{KL}(P \| Q) = \int_X p \ln \frac{p}{q} d\mu. \tag{6}$$

For the convex or concave functions $f(t) = t^\alpha$ we obtain the *Hellinger integrals* (e.g. [17])

$$H_\alpha(P,Q) = \int_X p^\alpha q^{1-\alpha} d\mu. \tag{7}$$

Those are related to the Rényi divergence of order α, $\alpha \neq 1$, introduced by Rényi [41] (for $\alpha > 0$) as

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \ln\left(\int_X p^\alpha q^{1-\alpha} d\mu\right) = \frac{1}{\alpha-1} \ln\left(H_\alpha(P,Q)\right). \tag{8}$$

The case $\alpha = 1$ is the relative entropy $D_{KL}(P\|Q)$.

3 f-divergences for convex bodies.

We will now consider f-divergences for convex bodies $K \in \mathcal{K}_0$. Let

$$p_K(x) = \frac{\kappa_K(x)}{\langle x, N_K(x)\rangle^n \; n|K^\circ|}, \quad q_K(x) = \frac{\langle x, N_K(x)\rangle}{n\;|K|}. \tag{9}$$

Usually, in the literature, the measures under consideration are probability measures. Therefore we have normalized the densities. Thus

$$P_K = p_K\;\mu_K \quad \text{and} \quad Q_K = q_K\;\mu_K \tag{10}$$

are measures on ∂K that are absolutely continuous with respect to μ_K. Q_K is a probability measure and P_K is one if K is in C^2_+.

Recall that the normalized cone measure cm_K on ∂K is defined as follows: For every measurable set $A \subseteq \partial K$

$$cm_K(A) = \frac{1}{|K|}\left|\left\{ta: \; a \in A, t \in [0,1]\right\}\right|. \tag{11}$$

The next proposition is well known. See e.g. [39] for a proof. It shows that the measures P_K and Q_K defined in (10) are the cone measures of K and K°. $N_K : \partial K \to S^{n-1}$, $x \to N_K(x)$ is the Gauss map.

Proposition 3.1. *Let K be a convex body in $\mathbb{R}^n$. Let P_K and Q_K be the probability measures on ∂K defined by (10). Then*

$$Q_K = cm_K,$$

or, equivalently, for every measurable subset A in ∂K $Q_K(A) = cm_K(A)$. If K is in addition in C^2_+, then

$$P_K = N_K^{-1} N_{K^\circ} cm_{K^\circ}$$

or, equivalently, for every measurable subset A in ∂K

$$P_K(A) = cm_{K^\circ}\left(N_{K^\circ}^{-1}\big(N_K(A)\big)\right). \tag{12}$$

It is in the sense (12) that we understand P_K to be the "cone measure" of K° and we write $P_K = cm_{K^\circ}$.

We now define the f-divergences of $K \in \mathcal{K}_0$. Note that $\langle x, N_K(x)\rangle > 0$ for all $x \in \partial K$ and therefore $\{x \in \partial K : q_K(x) = 0\} = \emptyset$. Hence, possibly also using our convention $0 \cdot \infty = 0$,

$$f^*(0)\ P_K\left(\{x \in \partial K : q_K(x) = 0\}\right) = 0.$$

Definition 3.2. *Let K be a convex body in $\mathcal{K}_0$ and let Let $f : (0,\infty) \to \mathbb{R}$ be a convex function. The f-divergence of K with respect to the cone measures P_K and Q_K is*

$$\begin{aligned} D_f(P_K, Q_K) &= \int_{\partial K} f\left(\frac{p_K}{q_K}\right) q_K d\mu_K \\ &= \int_{\partial K} f\left(\frac{|K|\kappa_K(x)}{|K^\circ|\langle x, N_K(x)\rangle^{n+1}}\right) \frac{\langle x, N_K(x)\rangle}{n|K|} d\mu_K. \end{aligned} \tag{13}$$

Remarks.

By (4) and (13)

$$\begin{aligned} D_f(Q_K, P_K) &= \int_{\partial K} f\left(\frac{q_K}{p_K}\right) p_K d\mu_K = D_{f^*}(P_K, Q_K) \\ &= \int_{\partial K} f^*\left(\frac{p_K}{q_K}\right) q_K d\mu_K \\ &= \int_{\partial K} f\left(\frac{|K^\circ|\langle x, N_K(x)\rangle^{n+1}}{|K|\kappa_K(x)}\right) \frac{\kappa_K(x)\ d\mu_K}{n|K^\circ|\langle x, N_K(x)\rangle^n}. \end{aligned} \tag{14}$$

f-divergences can also be expressed as integrals over S^{n-1},

$$D_f(P_K, Q_K) = \int_{S^{n-1}} f\left(\frac{|K|}{|K^\circ| f_K(u) h_K(u)^{n+1}}\right) \frac{h_K(u) f_K(u)}{n|K|} d\sigma \tag{15}$$

and

$$D_f(Q_K, P_K) \quad = \quad \int_{S^{n-1}} f\left(\frac{|K^\circ| f_K(u) h_K(u)^{n+1}}{|K|}\right) \frac{d\sigma_K}{n|K^\circ| h_K(u)^n}. \quad (16)$$

Examples.

If K is a polytope, the Gauss curvature κ_K of K is 0 a.e. on ∂K. Hence

$$D_f(P_K, Q_K) = f(0) \text{ and } D_f(Q_K, P_K) = f^*(0). \quad (17)$$

For every ellipsoid $\mathcal{E}$,

$$D_f(P_\mathcal{E}, Q_\mathcal{E}) = D_f(Q_\mathcal{E}, P_\mathcal{E}) = f(1) = f^*(1). \quad (18)$$

Denote by $Conv(0,\infty)$ the set of functions $\psi : (0,\infty) \to (0,\infty)$ such that ψ is convex, $\lim_{t\to 0} \psi(t) = \infty$, and $\lim_{t\to\infty} \psi(t) = 0$. For $\psi \in Conv(0,\infty)$, Ludwig [21] introduces the L_ψ affine surface area for a convex body K in $\mathbb{R}^n$

$$\Omega_\Psi(K) = \int_{\partial K} \psi\left(\frac{\kappa_K(x)}{\langle x, N_K(x)\rangle^{n+1}}\right) \langle x, N_K(x)\rangle d\mu_K. \quad (19)$$

Thus, L_ψ affine surface areas are special cases of (non-normalized) f-divergences for $f = \psi$.

For $\psi \in Conv(0,\infty)$, the $*$-adjoint function ψ^* is convex, $\lim_{t\to 0} \psi(t) = 0$, and $\lim_{t\to\infty} \psi(t) = \infty$. Thus ψ^* is an *Orlicz function* (see [18]), and gives rise to the corresponding *Orlicz-divergences* $D_{\psi^*}(P_K, Q_K)$ and $D_{\psi^*}(Q_K, P_K)$.

Let $p \leq 0$. Then the function $f : (0,\infty) \to (0,\infty)$, $f(t) = t^{\frac{p}{n+p}}$, is convex. The corresponding (non-normalized) f-divergence (which is also an Orlicz-divergence) is the L_p *affine surface area*, introduced by Lutwak [26] for $p > 1$ and by Schütt and Werner [47] for $p < 1, p \neq -n$. See also [12].

It was shown in [52] that all L_p affine surface areas are entropy powers of Rényi divergences.

For $p \geq 0$, the function $f : (0,\infty) \to (0,\infty)$, $f(t) = t^{\frac{p}{n+p}}$ is concave. The corresponding L_p affine surface areas $\int_{\partial K} \frac{\kappa_K^{\frac{p}{n+p}} d\mu_K}{\langle x, N_K(x)\rangle^{\frac{n(p-1)}{n+p}}}$ are examples of L_ϕ affine surface areas which were considered in [23] and [21]. Those, in turn are special cases of (non-normalized) f-divergences for *concave* functions f.

Let $f(t) = t\ln t$. Then the $*$-adjoint function is $f^*(t) = -\ln t$. The corresponding f-divergence is the *Kullback Leibler divergence* or *relative entropy* $D_{KL}(P_K\|Q_K)$ from P_K to Q_K

$$D_{KL}(P_K\|Q_K) = \int_{\partial K} \frac{\kappa_K(x)}{n|K^\circ|\langle x, N_K(x)\rangle^n} \ln\left(\frac{|K|\kappa_K(x)}{|K^\circ|\langle x, N_K(x)\rangle^{n+1}}\right) d\mu_K. \quad (20)$$

The relative entropy $D_{KL}(Q_K\|P_K)$ from Q_K to P_K is

$$\begin{aligned} D_{KL}(Q_K\|P_K) &= D_{f^*}(P_K, Q_K) & (21)\\ &= \int_{\partial K} \frac{\langle x, N_K(x)\rangle}{n|K|} \log\left(\frac{|K^\circ|\langle x, N_K(x)\rangle^{n+1}}{|K|\kappa_K(x)}\right) d\mu_K. & (22)\end{aligned}$$

Those were studied in detail in [39].

Equations (15) and (16) of the above remark lead us to define f-divergences for several convex bodies, or mixed f-divergences.

Let $K_1, \dots, K_n$ be convex bodies in $\mathcal{K}_0$. Let $u \in S^{n-1}$. For $1 \le i \le n$, define

$$p_{K_i}(u) = \frac{1}{n|K_i^\circ| h_{K_i}(u)}, \quad q_{K_i}(u) = \frac{f_{K_i}(u) h_{K_i}(u)}{n\ |K_i|}. \quad (23)$$

and measures on S^{n-1} by

$$P_{K_i} = p_{K_i}\ \sigma \quad \text{and} \quad Q_{K_i} = q_{K_i}\ \sigma. \quad (24)$$

Let $f_i : (0, \infty) \to \mathbb{R}$, $1 \le i \le n$, be convex functions. Then we define the *mixed f-divergences* for convex bodies $K_1, \dots, K_n$ in $\mathcal{K}_0$ by

Definition 3.3.

$$D_{f_1 \dots f_n}(P_{K_1} \times \dots \times P_{K_n}, Q_{K_1} \times \dots \times Q_{K_n}) = \int_{S^{n-1}} \prod_{i=1}^n \left[f_i\left(\frac{p_{K_i}}{q_{K_i}}\right) q_{K_i}\right]^{\frac{1}{n}} d\sigma$$

and

$$D_{f_1 \dots f_n}(Q_{K_1} \times \dots \times Q_{K_n}, P_{K_1} \times \dots \times P_{K_n}) = \int_{S^{n-1}} \prod_{i=1}^n \left[f_i\left(\frac{q_{K_i}}{p_{K_i}}\right) p_{K_i}\right]^{\frac{1}{n}} d\sigma.$$

Note that

$$D_{f_1^* \cdots f_n^*}(P_{K_1} \times \cdots \times P_{K_n}, Q_{K_1} \times \cdots \times Q_{K_n}) \\ = D_{f_1 \cdots f_n}(Q_{K_1} \times \cdots \times Q_{K_n}, P_{K_1} \times \cdots \times P_{K_n}).$$

Here, we concentrate on f-divergence for one convex body. Mixed f-divergences are treated similarly. We also refer to [55], where they have been investigated for functions in $Conv(0, \infty)$.

The observation (17) about polytopes holds more generally.

Proposition 3.4. *Let K be a convex body in $\mathcal{K}_0$ and let $f : (0, \infty) \to \mathbb{R}$ be a convex function. If K is such that $\mu_K(\{p_K > 0\}) = 0$, then*

$$D_f(P_K, Q_K) = f(0) \quad \text{and} \quad D_f(Q_K, P_K) = f^*(0).$$

Proof. $\mu_K(\{p_K > 0\}) = 0$ iff $Q_K(\{p_K > 0\}) = 0$. Hence the assumption implies that $Q_K(\{p_K = 0\}) = 1$. Therefore,

$$\begin{aligned} D_f(P_K, Q_K) &= \int_{\partial K} f\left(\frac{p_K}{q_K}\right) q_K d\mu_K \\ &= \int_{\{p_K > 0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu_K + \int_{\{p_K = 0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu_K \\ &= f(0). \end{aligned}$$

By (4), $D_f(Q_K, P_K) = D_{f^*}(P_K, Q_K) = f^*(0)$.

The next proposition complements the previous one. In view of (18) and (27), it corresponds to the affine isoperimetric inequality for f-divergences. It was proved in [17] in a different setting and in the special case of $f \in Conv(0, \infty)$ by Ludwig [21]. We include a proof for completeness.

Proposition 3.5. *Let K be a convex body in $\mathcal{K}_0$ and let $f : (0, \infty) \to \mathbb{R}$ be a convex function. If K is such that $\mu_K(\{p_K > 0\}) > 0$, then*

$$D_f(P_K, Q_K) \geq f\left(\frac{P_K(\{p_K > 0\})}{Q_K(\{p_K > 0\})}\right) Q_K(\{p_K > 0\}) + f(0)\, Q_K(\{p_K = 0\}) \tag{25}$$

and

$$D_f(Q_K, P_K) \geq f^*\left(\frac{P_K(\{p_K > 0\})}{Q_K(\{p_K > 0\})}\right) Q_K(\{p_K > 0\}) + f^*(0)\, Q_K(\{p_K = 0\}). \tag{26}$$

If K is in C^2_+, or if f is decreasing, then

$$D_f(P_K,Q_K) \geq f(1) \quad \textit{and} \quad D_f(Q_K,P_K) \geq f^*(1) = f(1). \tag{27}$$

Equality holds in (25) and (26) iff f is linear or K is an ellipsoid. If K is in C^2_+, equality holds in both inequalities (27) iff f is linear or K is an ellipsoid. If f is decreasing, equality holds in both inequalities (27) iff K is an ellipsoid.

Remark. It is possible for f to be deceasing and linear without having equality in (27). To see that, let $f(t) = at + b$, $a < 0$, $b > 0$. Then, for polytopes K (for which $\mu_K(\{p_K > 0\}) = 0$), $D_f(P_K,Q_K) = f(0) = b > f(1) = a + b$. But, also in the case when $0 < \mu_K(\{p_K > 0\}) < 1$, strict inequality may hold.

Indeed, let $\varepsilon > 0$ be sufficiently small and let $K = B^n_\infty(\varepsilon)$ be a "rounded" cube, where we have "rounded" the corners of the cube B^n_∞ with sidelength 2 centered at 0 by replacing each corner with εB^n_2 Euclidean balls. Then $D_f(P_K,Q_K) = b + a\ P_K(\{p_K > 0\}) > b + a = f(1)$.

Proof of Proposition 3.5. Let K be such that $\mu_K(\{p_K > 0\}) > 0$, which is equivalent to $Q_K(\{p_K > 0\}) > 0$. Then, by Jensen's inequality,

$$\begin{aligned} D_f(P_K,Q_K) &= Q_K(\{p_K > 0\}) \int_{\{p_K>0\}} f\left(\frac{p_K}{q_K}\right) \frac{q_K d\mu_K}{Q_K(\{p_K > 0\})} \\ &+ f(0)\ Q_K(\{p_K = 0\}) \\ &\geq Q_K(\{p_K > 0\})\, f\left(\frac{P_K(\{p_K > 0\})}{Q_K(\{p_K > 0\})}\right) + f(0) Q_K(\{p_K = 0\}). \end{aligned}$$

Inequality (26) follows by (4), as $D_f(Q_K,P_K) = D_{f^*}(P_K,Q_K)$.

If K is in C^2_+, $Q_K(\{p_K > 0\}) = 1$, $Q_K(\{p_K = 0\}) = 0$, $P_K(\{p_K > 0\}) = 1$ and $P_K(\{p_K = 0\}) = 0$. Thus we get that $D_f(P_K,Q_K) \geq f(1)$ and $D_f(Q_K,P_K) \geq f^*(1) = f(1)$.

If f is decreasing, then, by Jensen's inequality

$$D_f(P_K,Q_K) = \int_{\partial K} f\left(\frac{p_K}{q_K}\right) q_K d\mu_K \geq f\left(\int_{\partial K} p_K d\mu_K\right) \geq f(1).$$

The last inequality holds as $\int_{\partial K} p_K d\mu_K \leq 1$ and as f is decreasing.

Equality holds in Jensen's inequality iff either f is linear or $\frac{p_K}{q_K}$ is constant. Indeed, if $f(t) = at + b$, then

$$\begin{aligned} D_f(P_K, Q_K) &= \int_{\{p_K>0\}} \left(a\frac{p_K}{q_K} + b\right) q_K d\mu_K + f(0)\ Q_K\left(\{p_K = 0\}\right) \\ &= aP_K\left(\{p_K > 0\}\right) + f(0). \end{aligned}$$

If f is not linear, equality holds iff $\frac{p_K}{q_K} = c$, c a constant. As by assumption $\mu_K\left(\{p_K > 0\}\right) > 0$, $c \neq 0$. By a theorem of Petty [40], this holds iff K is an ellipsoid.

The next proposition can be found in [17] in a different setting. Again, we include a proof for completeness.

Proposition 3.6. *Let K be a convex body in $\mathcal{K}_0$ and let $f : (0, \infty) \to \mathbb{R}$ be a convex function. Then*

$$D_f(P_K, Q_K) \leq f(0)+f^*(0)+f(1)\Big[Q_K(\{0 < p_K \leq q_K\})+P_K(\{0 < q_K \leq p_K\})\Big]$$

and

$$D_f(Q_K, P_K) \leq f(0)+f^*(0)+f(1)\Big[Q_K(\{0 < p_K \leq q_K\})+P_K(\{0 < q_K \leq p_K\})\Big].$$

If f is decreasing, the inequalities reduce to $D_f(P_K, Q_K) \leq f(0)$ respectively, $D_f(Q_K, P_K) \leq f^*(0)$.

Proof. It is enough to prove the first inequality. The second one follows immediately form the first by (4).

$$
\begin{aligned}
D_f(P_K,Q_K) &= \int_{\partial K} f\left(\frac{p_K}{q_K}\right) q_K d\mu \\
&= \int_{\{p_K>0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu + f(0)\ Q_K(\{p_K=0\}) \\
&= f(0)\ Q_K(\{p_K=0\}) + \int_{\{0<p_K\}\cap\{f'\geq 0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu \\
&\qquad\qquad + \int_{\{0<p_K\}\cap\{f'\leq 0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu \\
&\leq f(0)\left[Q_K(\{p_K=0\}) + Q_K\left(\{p_K>0\}\cap\{f'\leq 0\}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_{\{0<p_K\leq q_K\}\cap\{f'\geq 0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu + \int_{\{0<q_K\leq p_K\}\cap\{f'\geq 0\}} f\left(\frac{p_K}{q_K}\right) q_K d\mu \\
&\leq f(0) + f(1)\ Q_K\left(\{0<p_K\leq q_K\}\cap\{f'\geq 0\}\right) \\
&+ \int_{\{0<q_K\leq p_K\}\cap\{f'\geq 0\}} f^*\left(\frac{q_K}{p_K}\right) p_K d\mu \\
&= f(0) + f(1)\ Q_K\left(\{0<p_K\leq q_K\}\cap\{f'\geq 0\}\right) \\
&+ \int_{\{0<q_K\leq p_K\}\cap\{f'\geq 0\}\cap\{(f^*)'\geq 0\}} f^*\left(\frac{q_K}{p_K}\right) p_K d\mu \\
&+ \int_{\{0<q_K\leq p_K\}\cap\{f'\geq 0\}\cap\{(f^*)'\leq 0\}} f^*\left(\frac{q_K}{p_K}\right) p_K d\mu \\
&\leq f(0) + f(1)\ Q_K\left(\{0<p_K\leq q_K\}\cap\{f'\geq 0\}\right) \\
&+ f^*(1)\ P_K\left(\{0<q_K\leq p_K\}\cap\{f'\geq 0\}\cap\{(f^*)'\geq 0\}\right) \\
&+ f^*(0)\ P_K\left(\{0<q_K\leq p_K\}\cap\{f'\geq 0\}\cap\{(f^*)'\leq 0\}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq f(0) + f^*(0)\ P_K\left(\{0<q_K\leq p_K\}\cap\{f'\geq 0\}\right) \\
&+ f(1)\Big[Q_K(\{0<p_K\leq q_K\}\cap\{f'\geq 0\}) + P_K(\{0<q_K\leq p_K\}\cap\{f'\geq 0\})\Big].
\end{aligned}
$$

It follows from the last expression that, if f is decreasing, the inequality reduces to $D_f(P_K,Q_K)\leq f(0)$.

The next proposition shows that f-divergences are $GL(n)$ invariant and that non-normalized f-divergences are $SL(n)$ invariant valuations. For functions in $Conv(0,\infty)$, this was proved by Ludwig [21].

For functions in $Conv(0,\infty)$ the expressions are also lower semicontinuous, as it was shown in [21]. However, this need not be the case anymore if we assume just convexity of f. Indeed, let $f(t) = t^2$ and let $K = B_2^n$ be the Euclidean unit ball. Let $(K_j)_{j\in\mathbb{N}}$ be a sequence of polytopes that converges to B_2^n. As observed above, $D_f(P_{K_j}, Q_{K_j}) = f(0) = 0$ for all j. But $D_f(P_{B_2^n}, Q_{B_2^n}) = f(1) = 1$.

Let $\tilde{P}_K = \frac{\kappa_K \mu_K}{\langle x, N_K(x)\rangle^n}$ and $\tilde{Q}_K = \langle x, N_K(x)\rangle \mu_K$. Then we will denote by $D_f(\tilde{P}_K, \tilde{Q}_K)$ and $D_f(\tilde{Q}_K, \tilde{P}_K)$ the non-normalized f-divergences. We will also use the following lemma from [47] for the proof of Proposition 3.8.

Lemma 3.7. *Let K be a convex body in $\mathcal{K}_0$. Let $h : \partial K \to \mathbb{R}$ be an integrable function, and $T : \mathbb{R}^n \to \mathbb{R}^n$ an invertible, linear map. Then*

$$\int_{\partial K} h(x) d\mu_K = |\det(T)|^{-1} \int_{\partial T(K)} \frac{f(T^{-1}(y))}{\|T^{-1t}(N_K(T^{-1}(y)))\|} d\mu_{T(K)}.$$

Proposition 3.8. *Let K be a convex body in $\mathcal{K}_0$ and let $f : (0,\infty) \to \mathbb{R}$ be a convex function. Then $D_f(P_K, Q_K)$ and $D_f(Q_K, P_K)$ are $GL(n)$ invariant and $D_f(\tilde{P}_K, \tilde{Q}_K)$ and $D_f(\tilde{Q}_K, \tilde{P}_K)$ are $SL(n)$ invariant valuations.*

Proof. We use (e.g. [47]) that

$$\langle T(x), N_{T(K)}(T(x))\rangle = \frac{\langle x, N_K(x)\rangle}{\|T^{-1t}(N_K(x))\|},$$

and

$$\kappa_K(x) = \|T^{-1t}(N_K(x))\|^{n+1} \det(T)^2 \kappa_{T(K)}(T(x))$$

and Lemma 3.7 to get that

$$\begin{aligned} D_f(P_K, Q_K) &= \int_{\partial K} f\left(\frac{p_K(x)}{q_K(x)}\right) q_K(x) d\mu(x) \\ &= \frac{1}{|\det(T)|} \int_{\partial T(K)} \frac{f\left(\frac{p_K(T^{-1}(y))}{q_K(T^{-1}(y))}\right) q_K(T^{-1}(y)) d\mu_{T(K)}}{\|T^{-1t}(N_K(T^{-1}(y)))\|} \\ &= D_f(P_{T(K)}, Q_{T(K)}). \end{aligned}$$

The formula for $D_f(Q_K, P_K)$ follows immediately from this one and (4). The $SL(n)$ invariance for the non-normalized f-divergences is shown in the same way.

Now we show that $D_f(\tilde{P}_K, \tilde{Q}_K)$ and $D_f(\tilde{Q}_K, \tilde{P}_K)$ are valuations, i.e. for convex bodies K and L in $\mathcal{K}_0$ such that $K \cup L \in \mathcal{K}_0$,

$$D_f(\tilde{P}_{K\cup L}, \tilde{Q}_{K\cup L}) + D_f(\tilde{P}_{K\cap L}, \tilde{Q}_{K\cap L}) = D_f(\tilde{P}_K, \tilde{Q}_K) + D_f(\tilde{P}_L, \tilde{Q}_L). \quad (28)$$

Again, it is enough to prove this formula and the one for $D_f(\tilde{Q}_K, \tilde{P}_K)$ follows with (4). To prove (28), we proceed as in Schütt [44]. For completeness, we include the argument. We decompose

$$\partial(K \cup L) = (\partial K \cap \partial L) \cup (\partial K \cap L^c) \cup (K^c \cap \partial L),$$

$$\partial(K \cap L) = (\partial K \cap \partial L) \cup (\partial K \cap \text{int} L) \cup (\text{int} K \cap \partial L),$$
$$\partial K = (\partial K \cap \partial L) \cup (\partial K \cap L^c) \cup (\partial K \cap \text{int} L),$$
$$\partial L = (\partial K \cap \partial L) \cup (\partial K^c \cap \partial L) \cup (\text{int} K \cap \partial L),$$

where all unions on the right hand side are disjoint. Note that for x such that the curvatures $\kappa_K(x)$, $\kappa_L(x)$, $\kappa_{K\cup L}(x)$ and $\kappa_{K\cap L}(x)$ exist,

$$\langle x, N_K(x)\rangle = \langle x, N_L(x)\rangle = \langle x, N_{K\cap L}(x)\rangle = \langle x, N_{K\cup L}(x)\rangle \tag{29}$$

and

$$\kappa_{K\cup L}(x) = \min\{\kappa_K(x), \kappa_L(x)\}, \quad \kappa_{K\cap L}(x) = \max\{\kappa_K(x), \kappa_L(x)\}. \tag{30}$$

To prove (28), we split the involved integral using the above decompositions and (29) and (30).

4 Geometric characterization of f-divergences.

In [52], geometric characterizations were proved for Rényi divergences. Now, we want to establish such geometric characterizations for f-divergences as well. We use the *surface body* [47] but the *illumination surface body* [54] or the *mean width body* [13] can also be used.

Let K be a convex body in $\mathbb{R}^n$. Let $g : \partial K \to \mathbb{R}$ be a nonnegative, integrable, function. Let $s \geq 0$.

The *surface body* $K_{g,s}$, introduced in [47], is the intersection of all closed half-spaces H^+ whose defining hyperplanes H cut off a set of $f\mu_K$-measure less than or equal to s from ∂K. More precisely,

$$K_{g,s} = \bigcap_{\int_{\partial K\cap H^-} g d\mu_K \leq s} H^+.$$

For $x \in \partial K$ and $s > 0$

$$x_s = [0, x] \cap \partial K_{g,s}.$$

The *minimal function* $M_g : \partial K \to \mathbb{R}$

$$M_g(x) = \inf_{0<s} \frac{\int_{\partial K \cap H^-(x_s, N_{K_{g,s}}(x_s))} g \, d\mu_K}{\mathrm{vol}_{n-1}\left(\partial K \cap H^-(x_s, N_{K_{g,s}}(x_s))\right)} \tag{31}$$

was introduced in [47]. $H(x,\xi)$ is the hyperplane through x and orthogonal to ξ. $H^-(x,\xi)$ is the closed halfspace containing the point $x+\xi$, $H^+(x,\xi)$ the other halfspace.

For $x \in \partial K$, we define $r(x)$ as the maximum of all real numbers ρ so that $B_2^n(x - \rho N_K(x), \rho) \subseteq K$. Then we formulate an integrability condition for the minimal function

$$\int_{\partial K} \frac{d\mu_K(x)}{(M_g(x))^{\frac{2}{n-1}} \, r(x)} < \infty. \tag{32}$$

The following theorem was proved in [47].

Theorem 4.1. *Let K be a convex body in $\mathbb{R}^n$. Suppose that $f : \partial K \to \mathbb{R}$ is an integrable, almost everywhere strictly positive function that satisfies the integrability condition (32). Then*

$$c_n \lim_{s\to 0} \frac{|K| - |K_{g,s}|}{s^{\frac{2}{n-1}}} = \int_{\partial K} \frac{\kappa_K^{\frac{1}{n-1}}}{g^{\frac{2}{n-1}}} d\mu_K,$$

where $c_n = 2|B_2^{n-1}|^{\frac{2}{n-1}}$.

Theorem 4.1 was used in [47] to give geometric interpretations of L_p affine surface area and in [52] to give geometric interpretations of Rényi divergences. Now we use this theorem to give geometric interpretations of f-divergence for cone measures of convex bodies.

For a convex function $f : (0,\infty) \to \mathbb{R}$, let $g_f, h_f : \partial K \to \mathbb{R}$ be defined as

$$g_f(x) = \left[n|K^\circ|n^n|K|^n \frac{p_K q_K}{\left(f\left(\frac{p_K}{q_K}\right)\right)^{n-1}} \right]^{\frac{1}{2}} \tag{33}$$

and

$$h_f(x) = g_{f^*}(x) = \left[n|K^\circ|n^n|K|^n \frac{q_K^n / p_K^{n-2}}{\left(f\left(\frac{p_K}{q_K}\right)\right)^{n-1}} \right]^{\frac{1}{2}}. \tag{34}$$

Corollary 4.2. *Let K be a convex body in $\mathcal{K}_0$ and let $f : (0, \infty) \to \mathbb{R}$ be convex. Let $g_f, h_f : \partial K \to \mathbb{R}$ be defined as in (33) and (34). If g_f and h_f are integrable, almost everywhere strictly positive functions that satisfy the integrability condition (32), then*

$$c_n \lim_{s \to 0} \frac{|K| - |K_{g_f,s}|}{s^{\frac{2}{n-1}}} = D_f(P_K, Q_K)$$

and

$$c_n \lim_{s \to 0} \frac{|K| - |K_{h_f,s}|}{s^{\frac{2}{n-1}}} = D_f(Q_K, P_K)$$

Proof. The proof of the corollary follows immediately from Theorem 4.1.

References

[1] M. S. Ali and D. Silvey, *A general class of coefficients of divergence of one distribution from another*, Journal of the Royal Statistical Society, Series B **28** (1966), 131-142.

[2] I. Csiszár, *Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten*, Publ. Math. Inst. Hungar. Acad. Sci. ser. A **8** (1963), 84-108.

[3] T. Cover and J. Thomas, *Elements of information theory*, second ed., Wiley-Interscience, (John Wiley and Sons), Hoboken, NJ, (2006).

[4] R. J. Gardner, *A positive answer to the Busemann-Petty problem in three dimensions*, Ann. of Math. (2) **140** (1994), 435-47.

[5] R. J. Gardner, *The dual Brunn-Minkowski theory for bounded Borel sets: Dual affine quermassintegrals and inequalities*, Adv. Math. **216** (2007), 358-386.

[6] R. J. Gardner, A. Koldobsky, and T. Schlumprecht, *An analytical solution to the Busemann-Petty problem on sections of convex bodies*, Ann. of Math. (2) **149** (1999), 691-703.

[7] R. J. Gardner and G. Zhang, *Affine inequalities and radial mean bodies.* Amer. J. Math. **120**, no.3, (1998), 505-528.

[8] E. Grinberg and G. Zhang, *Convolutions, transforms, and convex bodies*, Proc. London Math. Soc. (3) **78** (1999), 77-115.

[9] C. Haberl, *Blaschke valuations*, Amer. J. of Math., in press

[10] C. Haberl and F. Schuster, *General Lp affine isoperimetric inequalities.* J. Differential Geometry **83** (2009), 1-26.

[11] C. Haberl, E. Lutwak, D. Yang and G. Zhang, *The even Orlicz Minkowski problem*, Adv. Math. **224** (2010), 2485-2510

[12] D. Hug, *Curvature Relations and Affine Surface Area for a General Convex Body and its Polar.* Results in Mathematics **V. 29** 233-248, (1996).

[13] J. Jenkinson and E. Werner *Relative entropies for convex bodies*, to appear in Transactions of the AMS.

[14] D. Klain, *Star valuations and dual mixed volumes*, Adv. Math. **121** (1996), 80-101.

[15] D. Klain, *Invariant valuations on star-shaped sets*, Adv. Math. **125** (1997), 95-113.

[16] F. Liese and I. Vajda, *Convex Statistical Distances*, Leipzig, Germany, Teubner, (1987).

[17] F. Liese and I. Vajda, *On Divergences and Information in Statistics and Information Theory*, IEEE Transactions on Information Theory **52** (2006), 4394-4412.

[18] J. Lindenstrauss, L. Tzafriri, *Classical Banach spaces. I. Sequence spaces.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 92. Springer-Verlag, Berlin-New York, (1977).

[19] M. Ludwig, *Ellipsoids and matrix valued valuations*, Duke Math. J. **119** (2003), 159-188.

[20] M. Ludwig, *Minkowski areas and valuations*, J. Differential Geometry, **86** (2010), 133-162.

[21] M. Ludwig, *General affine surface areas*, Adv. Math., **224** (2010), 2346-2360.

[22] M. Ludwig and M. Reitzner, *A Characterization of Affine Surface Area*, Adv. Math. **147** (1999), 138-172.

[23] M. Ludwig and M. Reitzner, *A classification of $SL(n)$ invariant valuations.* Ann. of Math. **172** (2010), 1223-1271.

[24] E. Lutwak, *Dual mixed volumes,* Pacific J. Math. 58 (1975) 531-538.

[25] E. Lutwak, *The Brunn-Minkowski-Firey theory I : Mixed volumes and the Minkowski problem*, J. Differential Geom. **38** (1993), 131-150.

[26] E. Lutwak, *The Brunn-Minkowski-Firey theory II : Affine and geominimal surface areas*, Adv. Math. **118** (1996), 244-294.

[27] E. Lutwak, *Intersection bodies and dual mixed volumes*, Adv. Math. **71** (1988), 232-261.

[28] E. Lutwak, D. Yang and G. Zhang, *A new ellipsoid associated with convex bodies*, Duke Math. J. **104** (2000), 375-390.

[29] E. LUTWAK, D. YANG AND G. ZHANG, *Sharp Affine L_p Sobolev inequalities*, J. Differential Geometry **62** (2002), 17-38.

[30] E. LUTWAK, D. YANG AND G. ZHANG, *The Cramer–Rao inequality for star bodies*, Duke Math. J. **112** (2002), 59-81.

[31] E. LUTWAK, D. YANG AND G. ZHANG, *Volume inequalities for subspaces of L_p*, J. Differential Geometry **68** (2004), 159-184.

[32] E. LUTWAK, D. YANG AND G. ZHANG, *Moment-entropy inequalities*, Ann. Probab. **32** (2004), 757-774.

[33] E. LUTWAK, D. YANG AND G. ZHANG, *Cramer-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information*, IEEE Transactions on Information Theory **51** (2005), 473-478.

[34] E. LUTWAK AND G. ZHANG, *Blaschke-Santaló inequalities*, J. Differential Geom. **47** (1997), 1-16.

[35] M. MEYER AND E. WERNER, *The Santaló-regions of a convex body*, Transactions of the AMS **350**, no.11 (1998), 4569-4591.

[36] M. MEYER AND E. WERNER, *On the p-affine surface area*, Adv. Math. **152** (2000), 288-313.

[37] T. MORIMOTO, *Markov processes and the H-theorem*, J. Phys. Soc. Jap. **18** (1963), 328-331.

[38] F. NAZAROV, F. PETROV, D. RYABOGIN AND A. ZVAVITCH, *A remark on the Mahler conjecture: local minimality of the unit cube*, Duke Math. J. **154** (2010), 419-430.

[39] G. PAOURIS AND E. WERNER, *Relative entropy of cone measures and L_p-centroid bodies*, Proc. London Math. Soc. DOI 10.1112/plms/pdr030 (2011).

[40] C. PETTY, *Affine isoperimetric problems*, Discrete Geometry and Convexity. Ann. New York Acad. Sci. **441** (1985), 113-127.

[41] A. RÉNYI, *On measures of entropy and information*, Proceedings of the 4th Berkeley Symposium on Probability Theory and Mathematical Statistics, vol.1 (1961), 547-561.

[42] B. Rubin and G. Zhang, *Generalizations of the Busemann-Petty problem for sections of convex bodies*, J. Funct. Anal. **213** (2004), 473-501.

[43] F. Schuster, *Crofton measures and Minkowski valuations*, Duke Math. J. **154** (2010), 1-30.

[44] C. Schütt, *On the affine surface area*, Proc. Amer. Math. Soc. **118** (1990), 275-290.

[45] C. Schütt and E. Werner, *The convex floating body.* Math. Scand. **66** 275-290, (1990).

[46] C. Schütt and E. Werner, *Random polytopes of points chosen from the boundary of a convex body.* GAFA Seminar Notes, Lecture Notes in Mathematics **1807**, Springer-Verlag (2002), 241-422.

[47] C. Schütt and E. Werner, *Surface bodies and p-affine surface area.* Adv. Math. **187** (2004), 98-145.

[48] A. Stancu, *The Discrete Planar L_0-Minkowski Problem.* Adv. Math. **167** (2002), 160-174.

[49] A. Stancu, *On the number of solutions to the discrete two-dimensional L_0-Minkowski problem.* Adv. Math. **180** (2003), 290-323.

[50] E. Werner, *Illumination bodies and affine surface area,* Studia Math. **110** (1994), 257-269.

[51] E. Werner, *On L_p-affine surface areas*, Indiana Univ. Math. J. **56** No. 5 (2007), 2305-2324.

[52] E. Werner, *Rényi Divergence and L_p-affine surface area for convex bodies* Adv. Math. **230** (2012), 1040-1059.

[53] E. Werner and D. Ye, *New L_p affine isoperimetric inequalities*, Adv. Math. **218** (2008), no. 3, 762-780.

[54] E. Werner and D. Ye, *Inequalities for mixed p-affine surface area*, Math. Ann. **347** (2010), 703-737

[55] D. Ye, *Inequalities for general mixed affine surface areas*, J. London Math. Soc. **85** (2012), 101-120.

[56] G. Zhang, *Intersection bodies and Busemann-Petty inequalities in* $\mathbb{R}^4$, Ann. of Math. **140** (1994), 331-346.

[57] G. Zhang, *A positive answer to the Busemann-Petty problem in four dimensions*, Ann. of Math. **149** (1999), 535-543.

[58] G. Zhang, *New Affine Isoperimetric Inequalities*, ICCM 2007, Vol. II, 239-267.

Elisabeth Werner
Department of Mathematics
Case Western Reserve University
Cleveland, Ohio 44106, U. S. A.
elisabeth.werner@case.edu

Université de Lille 1
UFR de Mathématique
59655 Villeneuve d'Ascq, France