

# The dual Voronoi diagrams with respect to representational Bregman divergences

Frank Nielsen

*École Polytechnique / Sony Computer Science Laboratories*

*LIX/FRL*

*Palaiseau, France/ Tokyo, Japan*

*frank.nielsen@polytechnique.edu*

Richard Nock

*University of Antilles-Guyane*

*CEREGMIA*

*Martinique, France*

*Richard.Nock@martinique.univ-ag.fr*

**Abstract**—We present a *generalization* of Bregman Voronoi diagrams induced by a Bregman divergence acting on a representation function. Bregman divergences are canonical distortion measures of flat spaces induced by strictly convex and differentiable functions, called Bregman generators. Considering a representation function further allows us to conveniently embed the not necessarily flat source space into a dually flat space for which the dual Voronoi diagrams can be derived from an equivalent power affine diagram. We explain the fundamental dualities induced by the pair of Legendre convex conjugates coupled with a pair of conjugate representations. In particular, we show that Amari’s celebrated family of  $\alpha$ -divergences and Eguchi and Copas’s  $\beta$ -divergences are two cases of representational Bregman divergences that are often considered in information geometry. We report closed-form formula for their centroids and describe their dual Voronoi diagrams on the induced statistical manifolds.

**Keywords**-Voronoi diagrams, centroids, power diagrams, Bregman divergences,  $f$ -divergences,  $\alpha$ -divergences,  $\beta$ -divergences.

## I. INTRODUCTION AND PRIOR WORK

Let  $\mathcal{P} = \{P_1, \dots, P_n\} \in \mathcal{X}$  be a finite point set with respective vector coordinates  $\mathbf{p}_1, \dots, \mathbf{p}_n \in \mathbb{R}^d$  in a given Cartesian frame. The Voronoi diagram partitions the space according to *proximal regions*  $\text{vor}(P_i)$  defined with respect to a distance function  $D$  as

$$\text{vor}(P_i) = \{X \in \mathcal{X} \mid D(X, P_i) \leq D(X, P_j) \forall j \neq i\}.$$

The ordinary Voronoi diagram in Euclidean geometry is defined for the  $d$ -dimensional Euclidean distance

$$D(X, Y) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}.$$

Note that the ordinary Voronoi diagram is the same for *any* monotonously increasing function of the chosen distance function  $D(\cdot, \cdot)$ . Thus the ordinary Voronoi diagram can also be defined equivalently for the squared Euclidean distance  $D'(X, Y) = \sum_{i=1}^d (x_i - y_i)^2 = D^2(X, Y)$ .

Voronoi diagrams are fundamental data-structures of computational geometry [1]. They have been generalized by considering various types of objects (e.g., segments) and distances (e.g.,  $L_p$  metrics). Leibon and Letscher [2] studied the Voronoi diagram in Riemannian geometry. Onishi and

Imai [3], [4] considered the Voronoi diagram from the standpoint of information geometry [5]. Information geometry is primarily concerned with the intrinsic geometries of the statistical manifold of probability distributions; The fundamental distance measuring the separation of two probability distributions with densities  $p(x)$  and  $q(x)$  (interpreted as two points  $P$  and  $Q$  of the statistical manifold) is the Kullback-Leibler (KL) divergence:

$$\text{KL}(P||Q) = \text{KL}(p(x)||q(x)) = \int_x p(x) \log \frac{p(x)}{q(x)} dx \geq 0.$$

For discrete probability measures with mass functions  $\{p_i\}_i$  and  $\{q_i\}_i$  interpreted as  $d$ -dimensional points of  $\mathbb{R}_{+,*}^d$ , the discrete KL divergence is defined as

$$\text{KL}(P||Q) = \text{KL}(\{p_i\}_i||\{q_i\}_i) = \sum_{i=1}^d p_i \log \frac{p_i}{q_i} \geq 0.$$

Onishi and Imai [3], [4] studied the dual Voronoi diagrams induced by a general distortion measure (called a divergence that is not necessarily symmetric nor satisfying the triangle inequality) using the framework of information geometry [5]. This seminal study was further refined by Nielsen et al. [6], [7] under the framework of Bregman divergences [8] that induces all dually flat geometries, including the self-dual Euclidean geometry. Note that 2D Voronoi diagrams can be interactively rasterized [9] using the graphics processor (GPU).

In this paper, we further extend the study of Voronoi diagrams in information-theoretic spaces. We first present in Section II the notion of Bregman divergences coupled with a pair of conjugate representation functions. We show that non-flat geometries induced by  $\alpha$ -divergences [5] and  $\beta$ -divergences [10] can be cast into this Bregman representational framework. Section III describes the notion of generalized Bregman centroids, and reports closed form solutions for  $\alpha$ -means [11] and  $\beta$ -means [10]. Section IV generalizes the Bregman Voronoi study of Nielsen et al. [6], [7] to the case of dual representational Bregman Voronoi diagrams.

## II. GENERALIZING BREGMAN DIVERGENCES

### A. Bregman divergences and dually flat spaces

Let  $F$  be a strictly convex and differentiable function:  $F: \mathbb{R}^d \rightarrow \mathbb{R}$ . The Bregman divergence [8] between any two vector points  $\mathbf{p}$  and  $\mathbf{q}$  associated with the generator  $F$  is

$$B_F(\mathbf{p}||\mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p} - \mathbf{q}, \nabla F(\mathbf{q}) \rangle, \quad (1)$$

where  $\nabla F(\mathbf{x})$  denote the gradient of  $F$  at  $\mathbf{x} = [x_1 \dots x_d]^T$ . Choosing  $F(\mathbf{x}) = \mathbf{x}^T \mathbf{x} = \sum_{i=1}^d x_i^2$  yields the squared Euclidean distance:  $\|\mathbf{p} - \mathbf{q}\|^2$ . Setting  $F(\mathbf{x}) = \sum_{i=1}^d x_i \log x_i$  (the Shannon's negative entropy) yields the Kullback-Leibler divergence (KL):  $\sum_{i=1}^d p_i \log \frac{p_i}{q_i} + q_i - p_i$  also known as the relative entropy.

The divergence  $B_F$  can be written in a dual form using Legendre transformation. Let  $\mathbf{x}^* = \nabla F(\mathbf{x})$  be the one-to-one mapping defining another (non-linear) dual coordinate system in  $\mathbb{R}^d$ . The dual Bregman generator

$$F^*(\mathbf{x}^*) = \max_{\mathbf{x} \in \mathbb{R}^d} \{ \langle \mathbf{x}, \mathbf{x}^* \rangle - F(\mathbf{x}) \}$$

is convex in  $\mathbf{x}^*$ , and called the Legendre convex conjugate  $F^*$ . Strict convexity and differentiability is required in order to have the involution property  $F^{**} = F$ . Note that we have  $\nabla F^* = (\nabla F)^{-1}$ , see [6].

Thus the Bregman divergence originally defined by a convex generator  $F$  in Eq. 1, can be rewritten into a *canonical form* emphasizing on the convex conjugates  $F$  and  $F^*$  (with  $\mathbf{x}^* = \nabla F(\mathbf{x})$ ):

$$B_F(\mathbf{p}||\mathbf{q}) = F(\mathbf{p}) + F^*(\mathbf{q}^*) - \langle \mathbf{p}, \mathbf{q}^* \rangle. \quad (2)$$

It follows a dual divergence  $B_{F^*}$  such that  $B_F(\mathbf{p}||\mathbf{q}) = B_{F^*}(\mathbf{q}^*||\mathbf{p}^*)$ .

The two coordinate systems  $\mathbf{x}$  and  $\mathbf{x}^*$  define a dually flat structure in  $\mathbb{R}^d$  such that  $\mathbf{c}(\lambda) = (1 - \lambda)\mathbf{p} + \lambda\mathbf{q}$  is the  $F$ -geodesic and  $\mathbf{c}^*(\lambda) = (1 - \lambda)\mathbf{p}^* + \lambda\mathbf{q}^*$  is the dual  $F^*$ -geodesic passing through  $P$  and  $Q$ , two "straight" lines with respect to the dual coordinate system  $\mathbf{x}/\mathbf{x}^*$ . These two dual geodesics coincide only for the generalized quadratic generator  $F(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ , where  $\mathbf{A} \succ 0$  is a positive-definite matrix. In this case, we obtain the classical Euclidean geometry (self-dual geodesics for  $F(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{x} = F^*(\mathbf{x}^*)$  since  $\nabla F(\mathbf{x}) = \mathbf{x}$ ). See [5].

### B. Representation functions and generalized Bregman divergences

For sake of simplicity, consider *decomposable* Bregman divergences:

$$B_F(\mathbf{p}||\mathbf{q}) = \sum_{i=1}^d B_F(p_i||q_i),$$

where  $B_F(p||q)$  is a 1D Bregman divergence acting on scalars. With a slight abuse of notation, we write  $F(\mathbf{x}) =$

$\sum_{i=1}^d F(x_i)$  for a decomposable generator  $F$ . Consider a strictly monotonous representation function  $k(\cdot)$  that may introduce a non-linear coordinate system  $x_i = k(s_i)$  (and  $\mathbf{x} = k(\mathbf{s}) = [k(s_1) \dots k(s_d)]^T$ , where  $\mathbf{s} = [s_1 \dots s_d]^T$  denote the source coordinate system. Since  $k(\cdot)$  is strictly monotonous, the mapping is bijective and  $s_i = k^{-1}(x_i)$  (with  $\mathbf{s} = k^{-1}(\mathbf{x})$ ). We have the Bregman generator

$$U(\mathbf{x}) = \sum_{i=1}^d U(x_i) = \sum_{i=1}^d U(k(s_i)) = F(\mathbf{s})$$

with  $F = U \circ k$ . The dual 1D generator

$$U^*(x^*) = \max_x \{ x x^* - U(x) \}$$

induces the dual coordinate system  $x_i^* = U'(x_i)$ , where  $U'$  denotes the derivative of  $U$ . We note for short  $\nabla U(\mathbf{x}) = [U'(x_1) \dots U'(x_d)]^T$ . The dual separable generator is

$$U^*(\mathbf{x}^*) = \sum_{i=1}^d U^*(x_i^*),$$

and the canonical separable representational Bregman divergence follows

$$B_{U,k}(\mathbf{p}||\mathbf{q}) = U(k(\mathbf{p})) + U^*(k^*(\mathbf{q}^*)) - \langle k(\mathbf{p}), k^*(\mathbf{q}^*) \rangle, \quad (3)$$

with  $k^*(\mathbf{x}^*) = U'(k(\mathbf{x}))$ . This is in essence "quite" identical to Eq. 2 by setting  $F = U \circ k$ . However it turns out that although  $U$  is a strictly convex and differentiable function and  $k$  a strictly monotonous function,  $F = U \circ k$  may not be strictly convex (e.g., the  $\alpha$ -divergences described in Table I).

We can restate the representational Bregman divergence using a single coordinate system with a  $k$ -representation following Eq. 1

$$B_{U,k}(\mathbf{p}||\mathbf{q}) = U(k(\mathbf{p})) - U(k(\mathbf{q})) - \langle k(\mathbf{p}) - k(\mathbf{q}), \nabla U(k(\mathbf{q})) \rangle. \quad (4)$$

This is the Bregman divergence acting on the  $k$ -representation:

$$B_{U,k}(\mathbf{p}||\mathbf{q}) = B_U(k(\mathbf{p}), k(\mathbf{p})). \quad (5)$$

Note that for Bregman divergences of Eq. 1 with identity representation function  $k(\cdot)$ , the dual representation  $k^*(\mathbf{x}^*) = \nabla F(\mathbf{x})$  is not anymore the identity function. Thus function  $k(\cdot)$  can be seen as a (non-linear) embedding of the source space into a dually flat space. In particular, we show next that the renown  $\alpha$ -divergences [5] fit this generalization, although these divergences are characterized by constant curvature geometries [12]. It follows the dual representational divergence is  $B_{U^*,k^*}(\mathbf{p}^*||\mathbf{q}^*) = B_{U,k}(\mathbf{q}||\mathbf{p})$ .

### C. Information geometry, $\alpha$ -divergences and $\beta$ -divergences

In information geometry,  $\alpha$ -divergences [5] on positive arrays (unnormalized discrete probabilities) are defined for  $\alpha \in \mathbb{R}$  as

$$D_\alpha(\mathbf{p}||\mathbf{q}) = \begin{cases} \sum_{i=1}^d \frac{4}{1-\alpha^2} \left( \frac{1-\alpha}{2} p_i + \frac{1+\alpha}{2} q_i - p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}} \right) & \alpha \neq \pm 1 \\ \sum_{i=1}^d p_i \log \frac{p_i}{q_i} + q_i - p_i = \text{KL}(\mathbf{p}||\mathbf{q}) & \alpha = -1 \\ \sum_{i=1}^d q_i \log \frac{q_i}{p_i} + p_i - q_i = \text{KL}(\mathbf{q}||\mathbf{p}) & \alpha = 1 \end{cases} \quad (6)$$

Amari's  $\alpha$ -divergences are dual in the sense that  $D_\alpha(\mathbf{p}||\mathbf{q}) = D_{-\alpha}(\mathbf{q}||\mathbf{p})$ . These divergences are a special case of Csiszár  $f$ -divergences [13] associated to any convex function  $f$  satisfying  $f(1) = f'(1) = 0$

$$C_f(p||q) = \sum_{i=1}^d p_i f\left(\frac{q_i}{p_i}\right). \quad (7)$$

Indeed letting  $f$  be the parametric family of functions in Eq. 7 be

$$f_\alpha(x) = \frac{4}{1-\alpha^2} \left( \frac{1-\alpha}{2} + \frac{1+\alpha}{2} x - x^{\frac{1+\alpha}{2}} \right) \quad (8)$$

yields the formula of Eq. 6 of  $\alpha$ -divergences. These  $\alpha$ -divergences have proven useful in statistics [5]. Table I presents these  $\alpha$ -divergences as representational Bregman divergences. Note that  $F_\alpha = U_\alpha \circ k_\alpha = \frac{2}{1+\alpha}x$  and  $F_\alpha^* = F_{-\alpha} = U_{-\alpha} \circ k_{-\alpha} = \frac{2}{1-\alpha}x$ . These functions are *not* Legendre conjugates since they are not strictly convex. Indeed, these functions are only linear, and the Legendre transform is ill-defined. However, the functions  $U_\alpha$  and  $U_\alpha^*$  are proper Legendre convex conjugates. Thus the non-linear  $\alpha$ -embedding obtained by using the  $k_\alpha$ -representation allows one to embed the  $\alpha$ -geometry into a dually flat space. Note that  $k_\alpha$  is a strictly monotonous increasing function.

Another important class of parameterized divergences are the family of  $\beta$ -divergences that have proven handy for robust estimations in statistics [14]. The  $\beta$ -divergences [10] are defined on positive arrays as

$$D_\beta(\mathbf{p}||\mathbf{q}) = \begin{cases} \sum_{i=1}^d q_i \log \frac{q_i}{p_i} + p_i - q_i = \text{KL}(\mathbf{q}||\mathbf{p}) & \beta = 0 \\ \sum_{i=1}^d \frac{1}{\beta+1} (p_i^{\beta+1} - q_i^{\beta+1}) - \frac{1}{\beta} q_i (p_i^\beta - q_i^\beta) & \beta > 0 \end{cases} \quad (9)$$

$\beta$ -divergences are also representational Bregman divergences as shown in Table I (with  $U_0(x) = \exp x$ ). Note that  $F_\beta(x) = \frac{1}{\beta+1}x^{\beta+1}$  and  $F_\beta^*(x) = \frac{x^{\beta+1}-x}{\beta(\beta+1)}$  are degenerated

to linear functions for  $\beta = 0$ , and that  $k_\beta$  is a strictly monotonous increasing function.

*Lemma 1:* The  $\alpha$ -divergences and  $\beta$ -divergences are representational Bregman divergences. Their underlying geometries can be embedded into dually flat spaces.

### III. CENTROIDS AND BARYCENTERS

The centroids (and barycenters) with respect to representational Bregman divergences are generalized means [15], [11] that let intervene both the potential function and the representation function. The right-sided barycenter  $\mathbf{b}^R$  and the left-sided barycenter  $\mathbf{b}^L$  of  $n$  points  $\mathbf{p}_1, \dots, \mathbf{p}_n$  with associated weights  $w_1, \dots, w_n$  (such that  $\|\mathbf{w}\| = 1$  and all  $w_i \geq 0$ ) are the unique minimizers

$$\mathbf{b}^R = \arg \min_{\mathbf{c} \in \mathcal{X}} \sum_{i=1}^n w_i B_{U,k}(\mathbf{p}_i||\mathbf{c}), \quad (10)$$

$$\mathbf{b}^L = \arg \min_{\mathbf{c} \in \mathcal{X}} \sum_{i=1}^n w_i B_{U,k}(\mathbf{c}||\mathbf{p}_i). \quad (11)$$

It follows from [15] that the right-sided and left-sided barycenters are respectively a  $k$ -mean, and a  $\nabla F$ -mean (for strictly convex  $F = U \circ k$ ) or the  $k$ -representation of a  $\nabla U$ -mean (for degenerated  $F = U \circ k$ ):

$$\mathbf{b}^R = k^{-1} \left( \sum_i w_i k(\mathbf{p}_i) \right), \quad (12)$$

$$\mathbf{b}^L = k^{-1} \left( \nabla U^* \left( \sum_i w_i \nabla U(k(\mathbf{p}_i)) \right) \right) \quad (13)$$

*Proof:*

$$\begin{aligned} & \min_{\mathbf{c}} \frac{1}{n} \sum_i B_{U,k}(\mathbf{p}_i||\mathbf{c}) \\ \equiv & \min_{\mathbf{c}} \frac{1}{n} \sum_i U(k(\mathbf{p}_i)) - U(k(\mathbf{c})) \\ & - \sum_i \langle k(\mathbf{p}_i) - k(\mathbf{c}), \nabla U(k(\mathbf{c})) \rangle \\ \equiv & \min_{\mathbf{c}} -U(k(\mathbf{c})) - \left\langle \frac{1}{n} \sum_i k(\mathbf{p}_i) - k(\mathbf{c}), \nabla U(k(\mathbf{c})) \right\rangle \\ \equiv & \min_{\mathbf{c}} B_{U,k} \left( \frac{1}{n} \sum_i k(\mathbf{p}_i) || k(\mathbf{c}) \right) \geq 0 \end{aligned}$$

It follows that this is minimized for  $k(\mathbf{c}) = \frac{1}{n} \sum_i k(\mathbf{p}_i)$  since  $B_{U,k}(\mathbf{p}||\mathbf{q}) = 0$  iff.  $\mathbf{p} = \mathbf{q}$ . Since  $k$  is strictly monotonous, we get  $\mathbf{c} = k^{-1} \left( \frac{1}{n} \sum_i k(\mathbf{p}_i) \right)$ . ■

Note that  $k^{-1} \circ U^{-1} = (U \circ k)^{-1}$  so that  $\mathbf{b}^L$  is merely a  $(U \circ k)$ -mean in general. Again in case  $F$  is not strictly convex, we need to consider the  $U$ -mean on the  $k$ -representation (e.g.,  $\alpha$ -means). The symmetrized barycenter can be computed using the similar bisection search described in [15]. This allows us to easily get closed form solutions

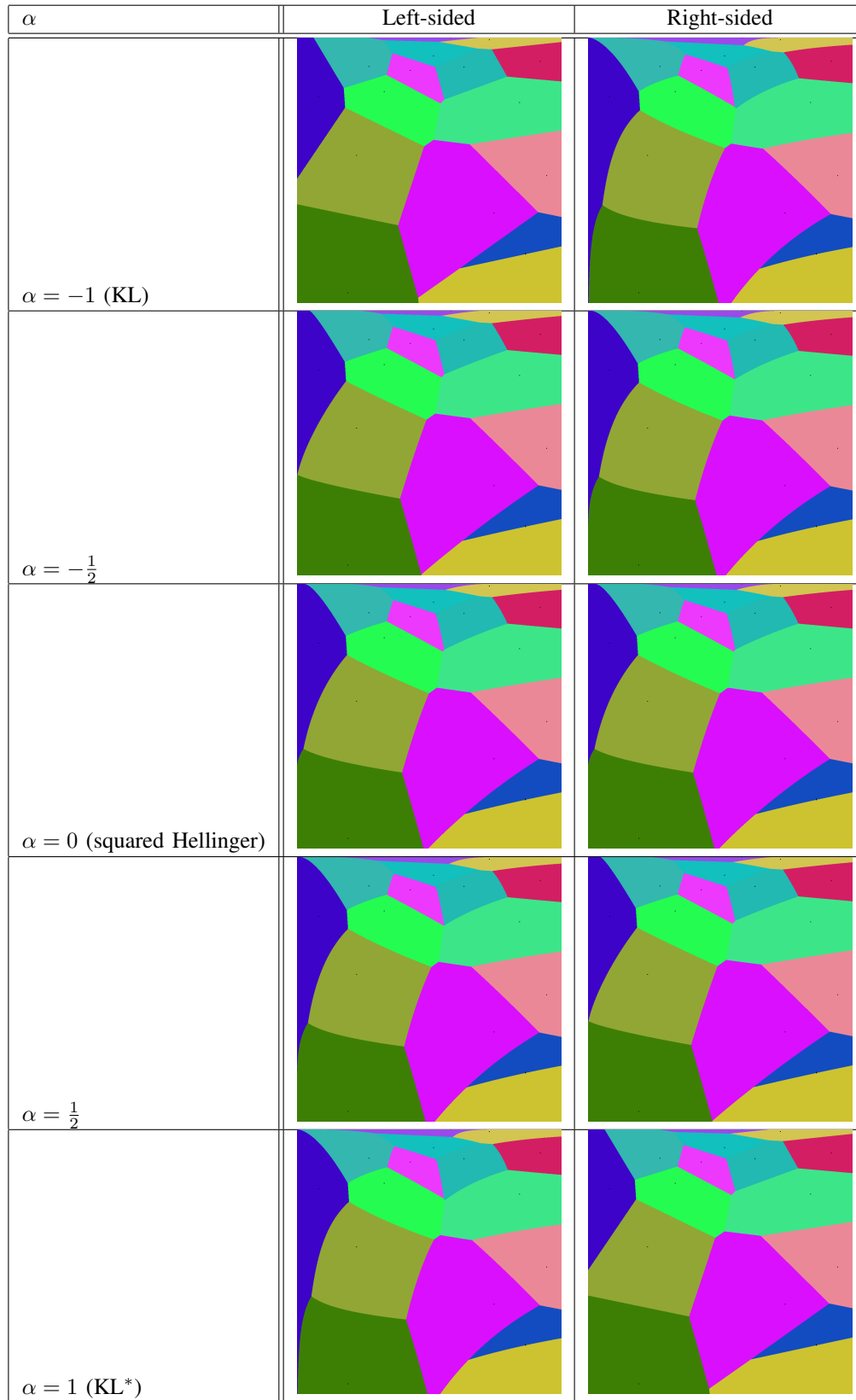


Figure 1. Voronoi diagrams for the family of  $\alpha$ -divergences. The right-sided (resp. left-sided) Voronoi diagram is affine for  $\alpha = 1$  (resp.  $\alpha = -1$ ).

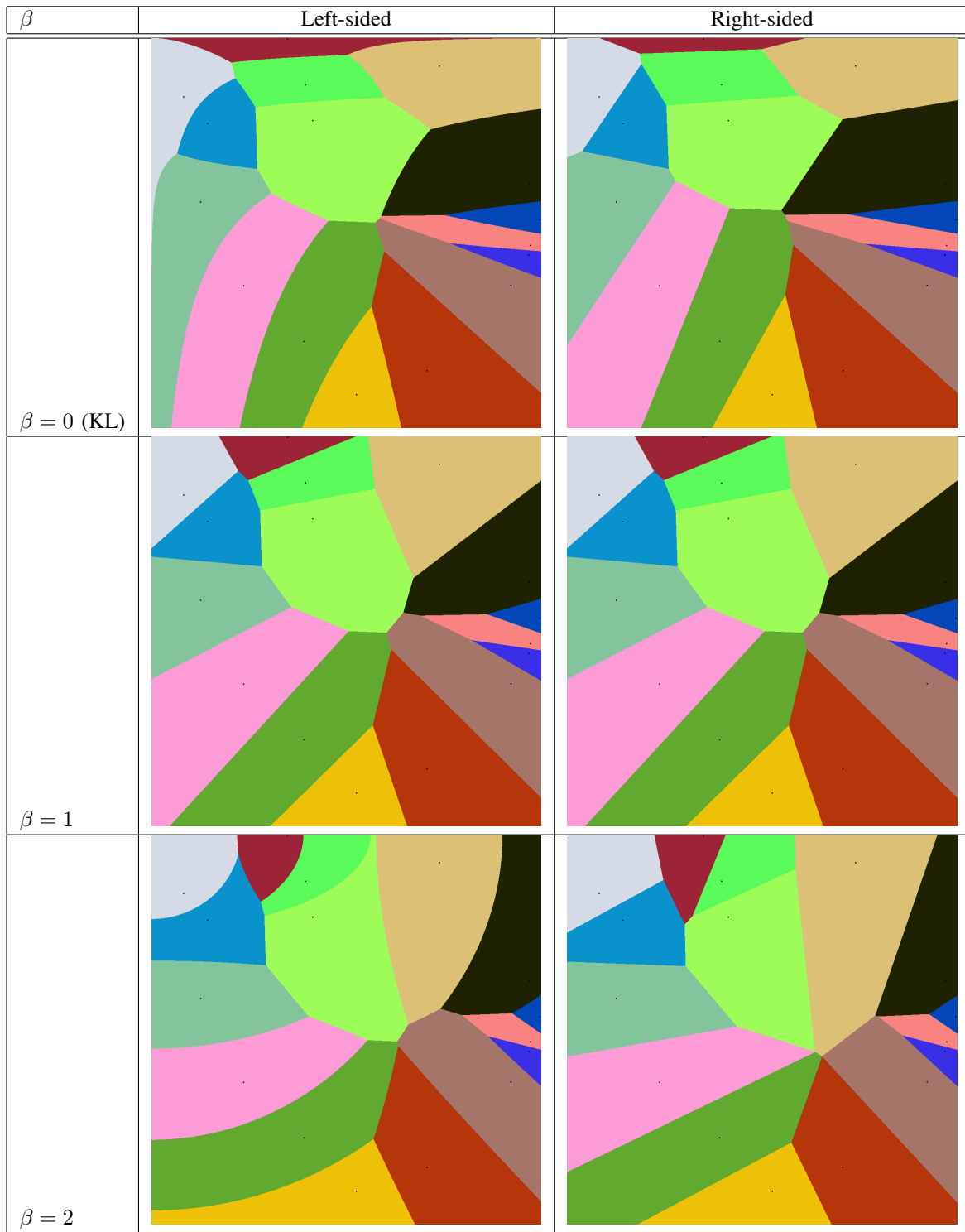


Figure 2. Voronoi diagrams for the family of  $\beta$ -divergences. The right-sided Voronoi diagram is affine.

Divergence	Convex conjugate functions	Representation functions
Bregman divergences $B_F, B_{F^*}$	$U$ $U' = (U^{*\prime})^{-1}$ $U^*$	$k(x) = x$ $k^*(x) = U'(k(x))$
$\alpha$ -divergences ( $\alpha \neq \pm 1$ ) $F_\alpha(x) = \frac{2}{1+\alpha}x$ $F_\alpha^*(x) = \frac{2}{1-\alpha}x$	$U_\alpha(x) = \frac{2}{1+\alpha}(\frac{1-\alpha}{2}x)^{\frac{2}{1-\alpha}}$ $U'_\alpha(x) = \frac{2}{1+\alpha}(\frac{1-\alpha}{2}x)^{\frac{1+\alpha}{1-\alpha}}$ $U_\alpha^*(x) = \frac{2}{1-\alpha}(\frac{1+\alpha}{2}x)^{\frac{2}{1+\alpha}} = U_{-\alpha}(x)$	$k_\alpha(x) = \frac{2}{1-\alpha}x^{\frac{1-\alpha}{2}}$ $k_\alpha^*(x) = \frac{2}{1+\alpha}x^{\frac{1+\alpha}{2}} = k_{-\alpha}(x)$
$\beta$ -divergences ( $\beta > 0$ ) $F_\beta(x) = \frac{1}{\beta+1}x^{\beta+1}$ $F_\beta^*(x) = \frac{x^{\beta+1}-x}{\beta(\beta+1)}$	$U_\beta(x) = \frac{1}{\beta+1}(1+\beta x)^{\frac{1+\beta}{\beta}}$ $U'_\beta(x) = (1+\beta x)^{\frac{1}{\beta}}$ $U_\beta^*(x) = \frac{x^{\beta+1}-x}{\beta(\beta+1)}$	$k_\beta(x) = \frac{x^\beta-1}{\beta}$ $k_\beta^*(x) = x$

Table I  
EXAMPLES OF REPRESENTATIONAL BREGMAN DIVERGENCES WITH THEIR CONJUGATE REPRESENTATION FUNCTIONS.

for the left/right sided  $\alpha$ -means and  $\beta$ -means, extending the seminal result of Amari on  $\alpha$ -means [11] (see Table II). For example, for  $\alpha \neq 1$ , we have  $k_\alpha(x) = \frac{2}{1-\alpha}(x^{\frac{1-\alpha}{2}} - 1)$  and  $k_\alpha^{-1}(y) = (1 + \frac{1-\alpha}{2}y)^{\frac{2}{1-\alpha}}$ . Applying the generalized right-sided centroid mean formula, we get

$$\mathbf{c} = k^{-1}\left(\frac{1}{n}\sum_{i=1}^n k(\mathbf{p}_i)\right) \quad (14)$$

$$= \left(1 + \frac{1-\alpha}{2} \frac{1}{n} \sum_{i=1}^n (\mathbf{p}_i^{\frac{1-\alpha}{2}} - 1)\right)^{\frac{2}{1-\alpha}} \quad (15)$$

$$= n^{-\frac{2}{1-\alpha}} \left(\sum_{i=1}^n \mathbf{p}_i^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}} \quad (16)$$

which corresponds to the  $\alpha$ -means of Amari [11]. Our method yields a short proof of Theorem 2 of [11].  $\alpha$ -means are thus generalized means that are written as  $k_\alpha^{-1}(\frac{1}{n}\sum_i k_\alpha(p_i))$ . Note that since  $D_\alpha(\mathbf{p}||\mathbf{q}) = D_{-\alpha}(\mathbf{q}||\mathbf{p})$ , we have the left-sided  $\alpha$ -mean is a right-sided  $-\alpha$ -mean, and vice-versa. Similarly, the right-sided  $\beta$ -means is a generalized mean obtained for  $k_\beta(x) = \frac{x^\beta-1}{\beta}$  and  $k_\beta^{-1}(x) = (1 + \beta x)^{\frac{1}{\beta}}$ . Note that we obtain the same right-sided means for the  $\alpha$ -means and  $\beta$ -means by taking  $\beta = \frac{1-\alpha}{2}$ . Having these generalized Bregman centroids defined, we can further extend<sup>1</sup> the Bregman  $k$ -means centroid-based clustering method [17] to that class of representational distortion measures, by first transforming the point set  $\mathcal{P}$  to its  $k$ -representation  $\mathcal{P}_k = \{k(\mathbf{p}_1), \dots, k(\mathbf{p}_n)\}$ .

<sup>1</sup>Furthermore, the remarkable bijection [16] of regular Bregman divergences with regular exponential families [16] extends to these embedded representations. (Straightforwardly, we get the notions of representational exponential families with the soft expectation-maximization clustering technique computed as a soft Bregman clustering on the embedded datasets.)

#### IV. REPRESENTATIONAL BREGMAN VORONOI DIAGRAMS

Since the distance is an asymmetric divergence ( $D(P||Q) \neq D(Q||P)$  with  $D = B_{U,k}$ ), we distinguish two left-sided and right-sided Voronoi diagrams defined by their Voronoi cells [6]:

$$\begin{aligned} \text{vor}_R(P_i) &= \{X \mid D(X||P_i) \leq D(X||P_j) \forall j \neq i\}, \\ \text{vor}_L(P_i) &= \{X \mid D(P_i||X) \leq D(P_j||X) \forall j \neq i\}, \end{aligned}$$

For a representational Bregman divergence, we have

$$D(k(\mathbf{p})||k(\mathbf{q})) = D^*(k^*(\mathbf{q}^*)||k^*(\mathbf{p}^*)),$$

so their the Voronoi diagrams have the same combinatorial complexity (mapping  $k(\cdot)$  is monotonous). Thus, we focus w.l.o.g. in the following on the right-sided Voronoi diagram.

##### A. Generalized Bregman Voronoi diagrams as lower envelopes

Voronoi diagrams can be obtained as minimization diagrams [1]:

$$\min_{i \in \{1, \dots, n\}} B_{U,k}(\mathbf{x}||\mathbf{p}_i).$$

This minimization diagram is equivalent to  $\min_i f_i(\mathbf{x})$  with

$$f_i(\mathbf{x}) = \langle k(\mathbf{p}_i) - k(\mathbf{x}), \nabla U(k(\mathbf{p}_i)) \rangle - U(k(\mathbf{p}_i)).$$

The functions  $f_i$ 's are linear in  $k(\mathbf{x})$  and denote hyperplanes. Thus by mapping the points  $\mathcal{P}$  to the point set  $\mathcal{P}_k$ , we obtain an affine minimization diagram that can be computed from the optimal half-space intersection algorithm of Chazelle [18]. Once the embedded Voronoi diagram is computed, we pull back this diagram by the strictly monotonous  $k^{-1}$  function.

*Lemma 2:* The Voronoi diagram of  $n$   $d$ -dimensional points with respect to a representational Bregman divergence has complexity  $O(n^{\lceil \frac{d}{2} \rceil})$ . It can be computed in  $O(n \log n + n^{\lceil \frac{d}{2} \rceil})$  time. It follows that the left/right sided Voronoi diagrams with respect to the  $\alpha$ -divergences or  $\beta$ -divergences have complexity  $O(n^{\lceil \frac{d}{2} \rceil})$ .

Means	Left-sided	Right-sided
Generic	$k^{-1} (\nabla U^* (\sum_{i=1}^n \frac{1}{n} \nabla U (k(\mathbf{p}_i))))$	$k^{-1} (\frac{1}{n} \sum_{i=1}^n k(\mathbf{p}_i))$
$\alpha$ -means ( $\alpha \neq \pm 1$ )	$n^{-\frac{2}{1+\alpha}} \left( \sum_{i=1}^n \mathbf{P}_i^{\frac{1+\alpha}{2}} \right)^{\frac{2}{1+\alpha}}$	$n^{-\frac{2}{1-\alpha}} \left( \sum_{i=1}^n \mathbf{P}_i^{\frac{1-\alpha}{2}} \right)^{\frac{2}{1-\alpha}}$
$\beta$ -means ( $\beta > 0$ )	$\frac{1}{n} \sum_{i=1}^n \mathbf{P}_i$	$n^{-\frac{1}{\beta}} \left( \sum_{i=1}^n \mathbf{P}_i^\beta \right)^{\frac{1}{\beta}}$

Table II  
LEFT-SIDED AND RIGHT-SIDED  $\alpha$ -MEANS AND  $\beta$ -MEANS.

To illustrate this theorem, consider the right-sided  $\alpha$ -bisectors

$$H_\alpha(\mathbf{p}, \mathbf{q}) : \{ \mathbf{x} \in \mathcal{X} \mid D_\alpha(\mathbf{p} \parallel \mathbf{x}) = D_\alpha(\mathbf{q} \parallel \mathbf{x}) \}$$

for  $\alpha \neq \pm 1$ . We get

$$H_\alpha(\mathbf{p}, \mathbf{q}) : \sum_i \frac{1-\alpha}{2} (p_i - q_i) + x^{\frac{1+\alpha}{2}} (q^{\frac{1-\alpha}{2}} - p^{\frac{1-\alpha}{2}}) = 0.$$

Let  $\mathbf{X} = [x_1^{\frac{1+\alpha}{2}} \dots x_d^{\frac{1+\alpha}{2}}]^T$ . Plugging  $X$  into the bisector equation (a special case of linearization), we get a hyperplane bisector:

$$H_\alpha(\mathbf{p}, \mathbf{q}) : \sum_i X_i (q^{\frac{1-\alpha}{2}} - p^{\frac{1-\alpha}{2}}) + \sum_i \frac{1-\alpha}{2} (p_i - q_i) = 0.$$

It follows that the right-sided  $\alpha$ -Voronoi diagram is affine in the  $k(x) = x^{\frac{1+\alpha}{2}}$  representation with complexity  $O(n^{\lceil \frac{d}{2} \rceil})$ . Indeed, we have

$$\begin{aligned} D(X \parallel P_i) = B_{U,k}(\mathbf{x} \parallel \mathbf{p}_i) &\leq D(X \parallel P_j) = B_{U,k}(\mathbf{x} \parallel \mathbf{p}_j) \\ \iff B_U(k(\mathbf{x}) \parallel k(\mathbf{p}_i)) &\leq B_U(k(\mathbf{x}) \parallel k(\mathbf{p}_j)). \end{aligned}$$

Note that right-sided  $\beta$ -Voronoi diagrams are affine for  $\beta > 0$ . Indeed, the  $\beta$ -bisector

$$H_\beta(\mathbf{p}, \mathbf{q}) : \{ \mathbf{x} \in \mathcal{X} \mid D_\beta(\mathbf{p} \parallel \mathbf{x}) = D_\beta(\mathbf{q} \parallel \mathbf{x}) \}$$

yields an equation linear in  $\mathbf{x}$ :

$$H_\beta(\mathbf{p}, \mathbf{q}) : \sum_{i=1}^d \frac{1}{\beta+1} (p_i^{\beta+1} - q_i^{\beta+1}) - \frac{1}{\beta} x_i (p_i^\beta - q_i^\beta) = 0.$$

### B. Generalized Bregman Voronoi diagrams from power diagrams

Let the power distance of a point  $\mathbf{x}$  to a Euclidean ball  $B = B(\mathbf{p}, r)$  defined by  $\|\mathbf{p} - \mathbf{x}\|^2 - r^2$ . The *power diagram* [19] of  $n$  balls  $B_i = B(\mathbf{p}_i, r_i)$ ,  $i = 1, \dots, n$  is the minimization diagram of the corresponding  $n$  functions  $D_i(\mathbf{x}) = \|\mathbf{p}_i - \mathbf{x}\|^2 - r_i^2$ . The power bisector of any two balls  $B(\mathbf{p}_i, r_i)$  and  $B(\mathbf{p}_j, r_j)$  is the radical hyperplane of equation

$$2\langle \mathbf{x}, \mathbf{p}_j - \mathbf{p}_i \rangle + \|\mathbf{p}_i\|^2 - \|\mathbf{p}_j\|^2 + r_j^2 - r_i^2 = 0.$$

Power diagrams are affine diagrams. Aurenhammer [19], [1] proved that *any* affine diagram is *identical* to the power diagram of a set of corresponding balls. Note that although some balls may have an empty cell in their power diagram, this case never occurs for representational Bregman divergences since  $B_{U,k}(\mathbf{p}_i \parallel \mathbf{p}_i) = 0$ . The mapping identifying the power bisector with the representational Bregman bisector is given by [6]

$$\mathbf{p}_i \rightarrow \nabla U(k(\mathbf{p}_i))$$

$$r_i = \langle U(k(\mathbf{p}_i)), U(k(\mathbf{p}_i)) \rangle + 2U(k(\mathbf{p}_i)) - \langle \mathbf{p}_i, U(k(\mathbf{p}_i)) \rangle.$$

## V. CONCLUDING REMARKS

We have generalized the study of Bregman Voronoi diagrams [6], [7] by introducing an extra representation function. The representation function can be interpreted as an embedding of the source space into a dually flat space. It followed that the dual Voronoi diagrams of  $n$   $d$ -dimensional points with respect to the  $\alpha$ -divergences [5] and  $\beta$ -divergences [10] can be constructed efficiently from power diagrams, and that their complexity is upper bounded by  $O(n^{\lceil \frac{d}{2} \rceil})$ . We can extend the notion of dual triangulations [7] and generalize the smallest enclosing ball algorithms [20], [21] to these representational Bregman divergences. Modern data analysis is focusing on recovering the topology, intrinsic dimension and underlying geometry of datasets. A challenging research axis is to learn from a dataset acquired with an unknown coordinate system both the underlying geometry (i.e., distance function) *and* the embedding (i.e., representation function), while reducing as efficiently as possible the risk of overfitting.

## REFERENCES

- [1] J.-D. Boissonnat and M. Yvinec, *Algorithmic Geometry*. New York, NY, USA: Cambridge University Press, 1998.
- [2] G. Leibon and D. Letscher, "Delaunay triangulations and Voronoi diagrams for Riemannian manifolds," in *SCG '00: Proceedings of the sixteenth annual symposium on Computational geometry*. New York, NY, USA: ACM, 2000, pp. 341–349.

- [3] K. Onishi and H. Imai, "Voronoi diagram in statistical parametric space by Kullback-Leibler divergence," in *SCG '97: Proceedings of the thirteenth annual symposium on Computational geometry*. New York, NY, USA: ACM, 1997, pp. 463–465.
- [4] —, "Voronoi diagram for the dually flat space by divergence," IPSJ SIG Notes, Tech. Rep. 42, 1997.
- [5] S. Amari and H. Nagaoka, *Methods of Information Geometry*, A. M. Society, Ed. Oxford University Press, 2000.
- [6] F. Nielsen, J.-D. Boissonnat, and R. Nock, "On Bregman Voronoi diagrams," in *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 746–755.
- [7] —, "Bregman Voronoi diagrams: Properties, algorithms and applications," *CoRR*, vol. abs/0709.2196, 2007.
- [8] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.
- [9] F. Nielsen, "An interactive tour of Voronoi diagrams on the GPU," in *ShaderX<sup>6</sup>: Advanced Rendering Techniques*. Charles River Media, 2008.
- [10] S. Eguchi and J. Copas, "A class of logistic-type discriminant functions," *Biometrika*, vol. 89, no. 1, pp. 1–22, March 2002. [Online]. Available: <http://dx.doi.org/10.1093/biomet/89.1.1>
- [11] S.-i. Amari, "Integration of stochastic models by minimizing  $\alpha$ -divergence," *Neural Comput.*, vol. 19, no. 10, pp. 2780–2796, 2007.
- [12] P. Vos, "Geometry of  $f$ -divergence," *Annals of the Institute of Statistical Mathematics*, vol. 43, no. 3, pp. 515–537, September 1991. [Online]. Available: <http://ideas.repec.org/a/spr/aistmt/v43y1991i3p515-537.html>
- [13] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Stat.*, vol. 19, pp. 2032–2066, 1991.
- [14] M. Mihoko and S. Eguchi, "Robust blind source separation by beta divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1859–1886, 2002.
- [15] F. Nielsen and R. Nock, "Sided and symmetrized Bregman centroids," *IEEE Transactions on Information Theory*, 2009.
- [16] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, 2005.
- [17] R. Nock, P. Luosto, and J. Kivinen, "Mixed Bregman clustering with approximation guarantees," in *ECML PKDD '08: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 154–169.
- [18] B. Chazelle, "An optimal convex hull algorithm in any fixed dimension," *Discrete Computational & Geometry*, vol. 10, pp. 377–409, 1993.
- [19] F. Aurenhammer, "Power diagrams: Properties, algorithms and applications," *SIAM Journal of Computing*, vol. 16, no. 1, pp. 78–96, 1987.
- [20] F. Nielsen and R. Nock, "On approximating the smallest enclosing bregman balls," in *SCG '06: Proceedings of the twenty-second annual symposium on Computational geometry*. New York, NY, USA: ACM, 2006, pp. 485–486.
- [21] —, "On the smallest enclosing information disk," *Inf. Process. Lett.*, vol. 105, no. 3, pp. 93–97, 2008.