On Bregman Voronoi Diagrams*

Frank Nielsen† Jean-Daniel Boissonnat‡ Richard Nock§

Abstract

The Voronoi diagram of a point set is a fundamental geometric structure that partitions the space into elementary regions of influence defining a discrete proximity graph and dually a well-shaped Delaunay triangulation. In this paper, we investigate a framework for defining and building the Voronoi diagrams for a broad class of distortion measures called Bregman divergences, that includes not only the traditional (squared) Euclidean distance, but also various divergence measures based on entropic functions. As a by-product, Bregman Voronoi diagrams allow one to define information-theoretic Voronoi diagrams in statistical parametric spaces based on the relative entropy of distributions. We show that for a given Bregman divergence, one can define several types of Voronoi diagrams related to each other by convex duality or embedding. Moreover, we can always compute them indirectly as power diagrams in primal or dual spaces, or directly after linearization in an extra-dimensional space as the projection of a Euclidean polytope. Finally, our paper proposes to generalize Bregman divergences to higher-order terms, called k-jet Bregman divergences, and touch upon their Voronoi diagrams.

1 Introduction and prior work

The *Voronoi diagram* $\mathrm{Vor}(\mathcal{S})$ of a set of n points $\mathcal{S} = \{\mathbf{p}_1, ..., \mathbf{p}_n\}$ of the d-dimensional Euclidean space $\mathbb{R}^d$ is defined as the *cell complex* (*ie.*, collection of cells of dimensions ranging from 0D vertices to $(d-1)$D facets) induced by its *Voronoi regions* $\{\mathrm{Vor}(\mathbf{p}_i)\}_{i\in\{1,..,n\}}$ and their subfaces, where $\mathrm{Vor}(\mathbf{p}_i)$ is the portion of points of $\mathbb{R}^d$ closer to $\mathbf{p}_i$ than to any other point of $\mathcal{S}$: $\mathrm{Vor}(\mathbf{p}_i) \stackrel{\mathrm{def}}{=} \{\mathbf{x} \in \mathbb{R}^d \mid ||\mathbf{p}_i\mathbf{x}|| \leq ||\mathbf{p}_j\mathbf{x}|| \ \forall\ \mathbf{p}_j \in \mathcal{S}\}$. Points $\{\mathbf{p}_i\}_i$ are called the *Voronoi sites* or *Voronoi generators*. Computational geometers have focused at first on *ordinary* Voronoi diagrams [3] by considering the L_2 norm $||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^d x_i^2}$ as the *distance function* (*ie.*, vector point sets $\mathcal{S}$ lying in the Euclidean space equipped with the Euclidean distance). The combinatorics and efficient algorithms for computing Voronoi diagrams in other *metric spaces* such as using the L_1 norm $||\mathbf{x}||_1 = \sum_{i=1}^d |x_i|$ (Manhattan distance) or L_∞ norm $||\mathbf{x}||_\infty = \max_{i\in\{1,...,d\}} x_i$ have later been reported [7, 3]. Klein [14] further presented an abstract framework for describing and computing the fundamental structures of abstract Voronoi diagrams.

In artificial intelligence, machine learning techniques also rely on geometric concepts for *building classifiers* in supervised problems (*eg.*, linear separators, oblique decision trees, etc.) or *clustering data* in unsupervised settings (*eg.*, `kmeans`, support vector clustering, etc.). However, the considered data sets $\mathcal{S}$ and their underlying spaces $\mathcal{X}$ are usually *not* metric spaces. That is, the distance function needs to be relaxed to a *(pseudo-)distance function* that does not necessarily satisfy anymore the *symmetry* nor the *triangle inequality*. To measure the *(dis)similarity* between any pair $(\mathbf{x}, \mathbf{y})$ of elements of $\mathcal{X}$, we introduce the notion of distortion between them by using a *divergence measure*. For example, in statistical spaces $\mathcal{X}$, vector points store in their coordinates parameters of the parametric distribution laws. Defining the "distance" between two such points (divergence between the corresponding distributions) is more delicate. Very few works have tackled an indepth study of Voronoi diagrams and their applications for such a kind of statistical spaces. Notable exceptions are the work of Onishi and Takayama [21] that focused on the Riemannian construction of the Voronoi diagram in the hyperbolic Poincaré half-plane[1] and the work of Leibon and Letscher [15] that focuses on Delaunay triangulations and Voronoi diagrams for Riemannian manifolds.

In this paper, we give a thorough treatment of Bregman Voronoi diagrams which elegantly *unifies* the ordinary Euclidean Voronoi diagram and statistical Voronoi diagrams. This is all the more important even for ordinary Voronoi diagrams as Euclidean point location of sites are usually observed in *noisy* environments (*e.g.*, imprecise point measures in computer vision experiments), and "noise" is often modelled by means of Normal distributions ("Gaussian noise").

To the best of our knowledge, the closest works with respect to statistical Voronoi diagrams to this Bregman

*`www.csl.sony.co.jp/person/nielsen/BregmanVoronoi/`

†Sony Computer Science Laboratories Inc., FRL, Japan.
‡INRIA Sophia-Antipolis, GEOMETRICA, France.
§University of Antilles-Guyane, CEREGMIA, France.

[1] The hyperbolic Voronoi diagram is also concisely described in [7], pp. 451–454.

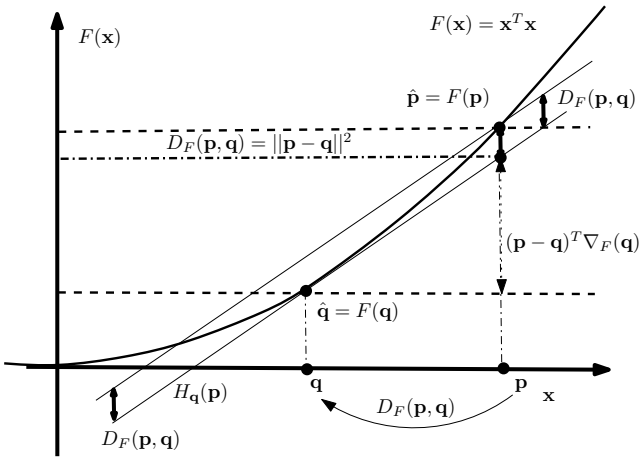

Figure 1: Visualizing the Bregman divergence for the squared Euclidean distance L_2^2.

Voronoi study is the 4-page short paper of Onishi and Imai [20] which relies on Kullback-Leibler divergence of dD multivariate Normal distributions to study combinatorics of their Voronoi diagrams, and subsequently the 2-page video paper of Sadakane et al. [24] which defines the divergence implied by a convex function and its conjugate, and present the Voronoi diagram with flavors of information geometry [1]. Our study of Bregman Voronoi diagrams generalizes these preliminary studies using a broader concept of divergence: Bregman divergences [8, 4] that do not rely *explicitly* on convex conjugates.

2 Bregman divergences

2.1 Definition Bregman[2] divergences [8] are parameterized families of distortion measures defined on a convex domain $\mathcal{X} \subseteq \mathbb{R}^d$ for strictly convex and differentiable functions F on $\mathrm{int}(\mathcal{X})$. Informally speaking, a Bregman divergence D_F is defined as the *tail* of a Taylor expansion of a strictly convex and differentiable function F: $D_F(\mathbf{p},\mathbf{q}) \stackrel{\mathrm{def}}{=} F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p}-\mathbf{q}, \boldsymbol{\nabla}_F(\mathbf{q})\rangle$,where $\boldsymbol{\nabla}_F$ denotes the gradient operator, and $\langle\cdot,\cdot\rangle$ the inner product (dot product). More precisely, we require F to be of Legendre-type (smooth). The strict convexity of F implies that $D_F(\mathbf{p},\mathbf{q}) \geq 0\ \forall \mathbf{p}\ \forall \mathbf{q}$ with $D_F(\mathbf{p},\mathbf{q}) = 0$ iff $\mathbf{p} = \mathbf{q}$ (positive definiteness). Bregman divergences unify the squared Euclidean distance (L_2^2 as depicted in Figure 1), several common distances such as Mahalanobis (popular in computer vision, also known as the generalized quadratic distance), and more importantly yield various information measures based on entropic functions such as Kullback-Leibler (Shannon entropy) or Itakura-Saito (Burg entropy, commonly used in sound processing).

Multivariate Bregman divergences D_F can be created from univariate Bregman functions coordinatewise as $F(\mathbf{x}) = \sum_{i=1}^d f_i(x_i)$ with $\boldsymbol{\nabla}_F = (f_1', ..., f_d')$ and $\boldsymbol{\nabla}_F{}^{-1} = ((f_1')^{-1}, ..., (f_d')^{-1})$ the reciprocal gradient: $\boldsymbol{\nabla}_F{}^{-1} \circ \boldsymbol{\nabla}_F = \boldsymbol{\nabla}_F \circ \boldsymbol{\nabla}_F{}^{-1} = \mathbf{I}$ (with $f_j \circ f_j^{-1} = f_j^{-1} \circ f_j = i\ \forall j \in \{1, ..., d\}$ the identity function $i(x) = x$). Common 1D Bregman divergences [4] are:

Squared Euclidean distance. $D_f(p,q) = (p-q)^2$ with $f = x^2$ defined on domain $\mathcal{X} = \mathbb{R}$ ($f' = 2x$ and $(f')^{-1} = \frac{x}{2}$).

Relative entropy (Kullback-Leibler divergence). $D_f(p,q) = p\log\frac{p}{q} - p + q$ with $f = x\log x$ the Shannon entropy on $\mathcal{X} = \mathbb{R}^+$ ($f' = \log x + 1$ and $(f')^{-1} = \exp(x-1)$).

Itakura-Saito divergence. $D_f(p,q) = \frac{p}{q} - \log\frac{p}{q} - 1$ with $f = -\log x$ the Burg entropy on $\mathcal{X} = \mathbb{R}^{+*}$ ($f' = -\frac{1}{x}$ and $(f')^{-1} = -\frac{1}{x}$).

Exponential divergence. $D_f(p,q) = \exp(p) - \exp(q) - (p-q)\exp(q)$ with $f = \exp x$ on $\mathcal{X} = \mathbb{R}$ ($f' = \exp x$ and $(f')^{-1} = \log x$).

Because the sum of convex functions is again a convex function, Bregman divergences can also be built by adding convex functions to create new generator functions.

2.2 Bregman divergences for statistical distributions A *parametric statistical space* Θ is a divergence space where vector points $\boldsymbol{\theta}$ represent the parameters of parametric distributions. The dimension d of the space coincides with the (finite) number of free parameters of the distribution laws. A large class of distribution families called the *exponential families* [1] admits the same *canonical* probabilistic distribution function: $p(x|\boldsymbol{\theta}) = h(x)Z(\boldsymbol{\theta})\exp\{\langle\boldsymbol{\theta}, \mathbf{f}(x)\rangle\}$ (with $\int_{-\infty}^{\infty} p(x|\boldsymbol{\theta})\mathrm{d}x = 1$). Exponential families include many famous distribution laws such as Poisson, Normal (univariate or multivariate Gaussian) and multinomial distributions. Vector $\mathbf{f}(x)$ represents the *sufficient statistics* and vector $\boldsymbol{\theta}$ stores the *natural parameters*. The *Kullback-Leibler divergence* (*relative entropy*) is an information-theoretic measure between two statistical distributions f and g defined as $\mathrm{KL}(f||g) \stackrel{\mathrm{def}}{=} \int_x f(x)\log\frac{f(x)}{g(x)}\mathrm{d}x$. The statistical measure is not necessarily symmetric nor does the triangle inequality hold. The Kullback-Leibler divergence of any two models of an exponential family with respective parameters $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_q$ is obtained from the Bregman divergence by choosing $F(\boldsymbol{\theta}) = -\log Z(\boldsymbol{\theta})$. This yields the

[2] Lev M. Bregman historically pionneered this notion in his seminal optimization work [8] (convex objective function to minimize under a set of linear constraints).

amount of information measure between two distributions of the *same* exponential family: $\mathrm{KL}(\boldsymbol{\theta}_p||\boldsymbol{\theta}_q) = D_F(\boldsymbol{\theta}_p, \boldsymbol{\theta}_q) = \langle(\boldsymbol{\theta}_p - \boldsymbol{\theta}_q), \boldsymbol{\theta}_p[\mathbf{f}]\rangle + \log \frac{Z(\boldsymbol{\theta}_q)}{Z(\boldsymbol{\theta}_p)}$ with $\boldsymbol{\theta}_p[\mathbf{f}] = [\int_x \frac{\mathbf{f}(x)}{Z(\boldsymbol{\theta}_p)} \exp\{\langle \boldsymbol{\theta}_p, \mathbf{f}(x)\rangle\}\mathrm{d}x]$. The coordinates of vector $\boldsymbol{\theta}_p[\mathbf{f}]$ are called the *expectation parameters*.

In summary, Bregman divergences include the squared Euclidean distance (Euclidean space) and the Kullback-Leibler divergence (statistical spaces). Note that in information geometry [1], statistical manifold properties are studied by the Riemannian geometry induced by the *Fischer metric* [1] of the considered exponential family. Bregman divergences have proven useful for a broad spectrum of applications in computer science and computational sciences. Fundamental applications of Bregman divergences are found in statistical inferences of generalized linear models.

3 Bregman Voronoi diagrams

3.1 First-type and second-type Bregman Voronoi diagrams Because Bregman divergences are not necessarily symmetric, we associate to each site $\mathbf{p}_i$ *two types* of functions, namely $D_i(\mathbf{x}) = D_F(\mathbf{x}, \mathbf{p}_i)$ or $D'_i(\mathbf{x}) = D_F(\mathbf{p}_i, \mathbf{x})$. The *minimization diagram* $\min_i D_i(\mathbf{x})$ of the D_i, $i = 1, \ldots, n$, is called the Bregman Voronoi diagram of the first type of $\mathcal{S}$, which we denote by $\mathrm{vor}_F(\mathcal{S})$. Similarly, the minimization diagram of the D'_i, $i = 1, \ldots, n$, is called the Bregman Voronoi diagram of the second type of $\mathcal{S}$, which we denote by $\mathrm{vor}'_F(\mathcal{S})$. A cell in $\mathrm{vor}_F(\mathcal{S})$ is associated to each site $\mathbf{p}_i$ and is defined as $\mathrm{vor}_F(\mathbf{p}_i) \stackrel{\mathrm{def}}{=} \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x}, \mathbf{p}_i) \leq D_F(\mathbf{x}, \mathbf{p}_j)\ \forall \mathbf{p}_j \in \mathcal{S}\}$. A cell in $\mathrm{vor}'_F(\mathcal{S})$ is associated to each site $\mathbf{p}_i$ and is equivalently defined as above with permuted divergence arguments: $\mathrm{vor}'_F(\mathbf{p}_i) \stackrel{\mathrm{def}}{=} \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{p}_i, \mathbf{x}) \leq D_F(\mathbf{p}_j, \mathbf{x})\ \forall \mathbf{p}_j \in \mathcal{S}\}$. Figure 2 illustrates the Bregman Voronoi diagrams of point sets for the Kullback-Leibler (relative entropy) and Itakura-Saito divergences. The ordinary Euclidean Voronoi diagram is a Bregman Voronoi diagram since $\mathrm{Vor}(\mathcal{S}) = \mathrm{vor}_F(\mathcal{S}) = \mathrm{vor}'_F(\mathcal{S})$ for $F(\mathbf{x}) = \sum_{i=1}^d x_i^2$, and in fact, for any strictly monotonically increasing family of distance functions.

Let $H(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x}, \mathbf{p}) = D_F(\mathbf{x}, \mathbf{q})\}$ be the Bregman bisector of first-type. Similarly, define the second-type bisector as $H'(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{p}, \mathbf{x}) = D_F(\mathbf{q}, \mathbf{x})\}$. These bisectors are matching hyperplanes in the Euclidean case; In the general case, a neat characterization follows from the Legendre-Fenchel convex duality. Let F^* be the Legendre conjugate of F obtained by the Legendre-Fenchel transformation $F^*(\mathbf{x}) = \sup\{\langle \mathbf{x}, \mathbf{y}\rangle - F(\mathbf{y}) \mid \mathbf{y} \in \mathcal{X}\} = \langle \mathbf{x}, {\boldsymbol{\nabla}_F}^{-1}(\mathbf{x})\rangle - F({\boldsymbol{\nabla}_F}^{-1}(\mathbf{x}))$. This transformation yields a *symmetric one-to-one mapping*, with $(F^*)^* = F$. Moreover,

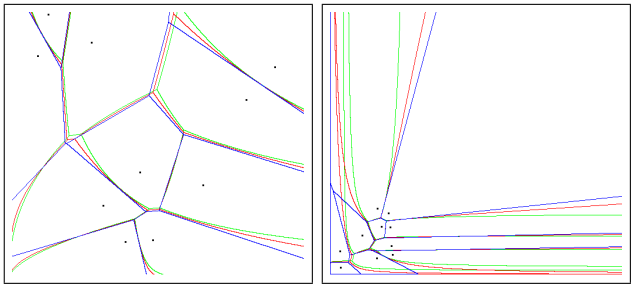

Figure 2: First-type affine Bregman Voronoi diagram (blue), second-type Bregman Voronoi diagram (green) and symmetrized Bregman Voronoi diagram (red) for the Kullback-Leibler relative entropy and Itakura-Saito divergences.

since F^* is strictly convex and differentiable (provided strictly convex and differentiable F), we can associate a *dual Bregman divergence* D_{F^*} to D_F using the conjugate function F^*, and we have the following equality: $D_F(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) + F^*(\mathbf{y}') - \langle \mathbf{x}, \mathbf{y}'\rangle = D_{F^*}(\mathbf{y}', \mathbf{x}')$ with $\mathbf{x}' = \boldsymbol{\nabla}_F(\mathbf{x})$ and $\mathbf{y}' = \boldsymbol{\nabla}_F(\mathbf{y})$. We note $\boldsymbol{\nabla}_F\mathcal{X} = \{\mathbf{x}' = \boldsymbol{\nabla}_F(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}$, and $\mathcal{S}_{\boldsymbol{\nabla}_F} = \{\boldsymbol{\nabla}_F(\mathbf{p}) \mid \mathbf{p} \in \mathcal{S}\}$ the gradient point set. With the help of these tools, we obtain the following lemma (see Fig. 3):

Lemma 3.1. *The first-type Bregman bisector of* $\mathbf{p}$ *and* $\mathbf{q}$ *for divergence* D_F *is a hyperplane in* $\mathcal{X}$*, of equation:* $H(\mathbf{p}, \mathbf{q}) : \langle \mathbf{x}, \mathbf{p}' - \mathbf{q}'\rangle + F(\mathbf{p}) - \langle \mathbf{p}, \mathbf{p}'\rangle - F(\mathbf{q}) + \langle \mathbf{q}, \mathbf{q}'\rangle = 0$ $(\star)$*. The second-type Bregman bisector* (H^*) *of* $\mathbf{p}$ *and* $\mathbf{q}$ *is the reciprocal of the first-type Bregman bisector for the gradient point set* $\mathcal{S}_{\boldsymbol{\nabla}_F}$*, and dual divergence* D_{F^*}*. As such, its image by* $\boldsymbol{\nabla}_F$ *is always a hyperplane in* $\boldsymbol{\nabla}_F\mathcal{X}$*.*

Proof. A point $\mathbf{x}$ belongs to the first-type bisector $H(\mathbf{p}, \mathbf{q})$ iff $D_F(\mathbf{x}, \mathbf{p}) = D_F(\mathbf{x}, \mathbf{q})$. Simplifying for $\mathbf{x}$ brings $(\star)$, a hyperplane since the equation may be shortened as $\langle \mathbf{x}, \mathbf{d_{pq}}\rangle + k_{\mathbf{pq}} = 0$, with $\mathbf{d_{pq}} = \mathbf{p}' - \mathbf{q}'$ a vector and $k_{\mathbf{pq}} = F(\mathbf{p}) - \langle \mathbf{p}, \mathbf{p}'\rangle - F(\mathbf{q}) + \langle \mathbf{q}, \mathbf{q}'\rangle$ a constant. Now, for the second-type Bregman bisector, remark that $D_F(\mathbf{p}, \mathbf{x}) = D_F(\mathbf{q}, \mathbf{x})$ iff $D_{F^*}(\mathbf{x}', \mathbf{p}') = D_{F^*}(\mathbf{x}', \mathbf{q}')$. Thus, the second-type Bregman bisector is the reciprocal of the first-type Bregman bisector for the gradient point set $\mathcal{S}_{\boldsymbol{\nabla}_F}$, computed using the dual divergence D_{F^*}. Note that the equation of the second-type bisector is $H^* : \langle \mathbf{x}', \mathbf{q} - \mathbf{p}\rangle + F(\mathbf{p}) - F(\mathbf{q}) = 0$. That is, linear in the gradient space. (Therefore, we do not require to compute the Legendre conjugate F^* of F to describe the second-type bisector.)

Corollary 3.1. *The linear equation of the first-type bisector for the Kullback-Leibler divergence associated with the exponential family is:* $H(\boldsymbol{\theta}_{\mathbf{x}}[\mathbf{f}]) :$

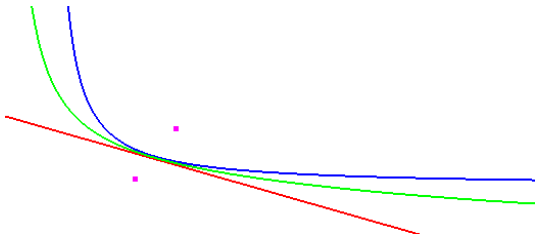

Figure 3: Three Itakura-Saito bisectors: first-type (red), second-type (blue) and third-type (green). The first-type Bregman bisector is *always* a linear separator. The second-type bisector is curved in the primal space but linear in the dual Legendre-Fenchel space $\boldsymbol{\nabla}_F\mathcal{X}$.

$\langle(\boldsymbol{\theta}_p - \boldsymbol{\theta}_q), \boldsymbol{\theta}_{\mathbf{x}}[\mathbf{f}]\rangle + \log \frac{Z(\boldsymbol{\theta}_p)}{Z(\boldsymbol{\theta}_q)} = 0$ *(linear in the expectation parameter variable* $\boldsymbol{\theta}_{\mathbf{x}}[\mathbf{f}]$*). The second-type bisector is given by* $H' : \langle\boldsymbol{\theta}_{\mathbf{x}}, \boldsymbol{\theta}_{\mathbf{q}}[\mathbf{f}] - \boldsymbol{\theta}_{\mathbf{p}}[\mathbf{f}]\rangle + \langle\boldsymbol{\theta}_{\mathbf{p}}, \boldsymbol{\theta}_{\mathbf{p}}[\mathbf{f}]\rangle - \langle\boldsymbol{\theta}_{\mathbf{q}}, \boldsymbol{\theta}_{\mathbf{q}}[\mathbf{f}]\rangle + \log \frac{Z(\boldsymbol{\theta}_{\mathbf{q}})}{Z(\boldsymbol{\theta}_{\mathbf{p}})} = 0$ *(linear in the natural parameter variable* $\boldsymbol{\theta}_{\mathbf{x}}$*).*

The construction duality unveiled in Lemmata 3.1 and 3.3 for bisectors also hold for Bregman Voronoi diagrams, as follows:

LEMMA 3.2. *The first-type Bregman Voronoi diagram is an affine diagram. The second-type curved Bregman Voronoi diagram is dual to the first-type Bregman Voronoi diagram of the gradient point set* $\mathcal{S}_{\boldsymbol{\nabla}_F}$*, and vice-versa. That is,* $\mathrm{vor}'_F(\mathcal{S}) \equiv \mathrm{vor}_{F^*}(\mathcal{S}_{\boldsymbol{\nabla}_F})$ *and* $\mathrm{vor}_F(\mathcal{S}) \equiv \mathrm{vor}'_{F^*}(\mathcal{S}_{\boldsymbol{\nabla}_F})$.

Proof. Let us write the definition of the first-type Bregman Voronoi region of site $\mathbf{p}_i$: $\mathrm{vor}_F(\mathbf{p}_i) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x}, \mathbf{p}_i) \leq D_F(\mathbf{x}, \mathbf{p}_j)\ \forall \mathbf{p}_j \in \mathcal{S}\}$. Introducing the dual divergence equality: $D_F(\mathbf{x}, \mathbf{y}) = D_{F^*}(\mathbf{y}', \mathbf{x}')$, we get $\mathrm{vor}_F(\mathbf{p}_i) = \{\mathbf{x}' \in \mathcal{X}_{\boldsymbol{\nabla}_F} \mid D_{F^*}(\mathbf{p}'_i, \mathbf{x}') \leq D_{F^*}(\mathbf{p}'_j, \mathbf{x}')\ \forall \mathbf{p}'_j \in \mathcal{S}_{\boldsymbol{\nabla}_F}\} = \mathrm{vor}'_{F^*}(\mathbf{p}'_i)$. That is: $\mathrm{vor}_F(\mathcal{S}) \equiv \mathrm{vor}'_{F^*}(\mathcal{S}_{\boldsymbol{\nabla}_F})$ (we omit the "dual" proof $\mathrm{vor}'_F(\mathcal{S}) \equiv \mathrm{vor}_{F^*}(\mathcal{S}_{\boldsymbol{\nabla}_F})$ that follows the same path).

Hence, to construct the second-type curved Bregman Voronoi diagram, we first convert the point set to its gradient set $\mathcal{S}_{\boldsymbol{\nabla}_F}$, and consider the affine Bregman Voronoi diagram using the dual Bregman divergence. Then we *convert back* the dual Bregman Voronoi diagram using the reciprocal gradient function ${\boldsymbol{\nabla}_F}^{-1}$. Building (first- or second-type) Bregman Voronoi diagrams thus reduces to constructing (primal or dual) affine Voronoi diagrams.

Before closing this section, we pursue the characterization of Bregman bisectors. Define $\mathrm{l}(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} : \exists\lambda \in \mathbb{R}, \mathbf{x} = \lambda\mathbf{p} + (1-\lambda)\mathbf{q}\}$ and $\mathrm{c}(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} : \exists\lambda \in \mathbb{R}, \mathbf{x}' = \lambda\mathbf{p}' + (1-\lambda)\mathbf{q}'\}$. In the Euclidean case (choosing $f(x) = \frac{x^2}{2}$), $\mathrm{l}(\mathbf{p}, \mathbf{q})$ and $\mathrm{c}(\mathbf{p}, \mathbf{q})$ are identical; they are orthogonal to bisector H in the Euclidean projections of $\mathbf{p}$ and $\mathbf{q}$ on H. For Bregman divergences, these two lines $\mathrm{l}(\mathbf{p}, \mathbf{q})$ and $\mathrm{c}(\mathbf{p}, \mathbf{q})$ appear to be different, yet we show that they enjoy a similar relationship as the one for the first and second-type Bregman bisectors (Lemma 3.1). First, define the *Bregman orthogonal projection* on H for some $\mathbf{x} \in \mathcal{X}$ as $\boldsymbol{\pi}_{\mathbf{x}} = \arg\min_{\mathbf{c}\in H} D_F(\mathbf{c}, \mathbf{x})$. This projection is *unique* and $(\mathbf{c}, \boldsymbol{\pi}_{\mathbf{x}}, \mathbf{x})$ satisfies the *generalized Pythagorean theorem* [13]: $D_F(\mathbf{c}, \boldsymbol{\pi}_{\mathbf{x}}) + D_F(\boldsymbol{\pi}_{\mathbf{x}}, \mathbf{x}) = D_F(\mathbf{c}, \mathbf{x}), \forall \mathbf{c} \in H$. As such, it generalizes the conventional Euclidean orthogonal projection, and we still have $\boldsymbol{\pi}_{\mathbf{p}} = \boldsymbol{\pi}_{\mathbf{q}}$. Otherwise indeed, we would have $D_F(\boldsymbol{\pi}_{\mathbf{q}}, \mathbf{p}) = D_F(\boldsymbol{\pi}_{\mathbf{q}}, \boldsymbol{\pi}_{\mathbf{p}}) + D_F(\boldsymbol{\pi}_{\mathbf{p}}, \mathbf{p}) > D_F(\boldsymbol{\pi}_{\mathbf{p}}, \mathbf{p})$; since $D_F(\boldsymbol{\pi}_{\mathbf{q}}, \mathbf{p}) = D_F(\boldsymbol{\pi}_{\mathbf{q}}, \mathbf{q})$ and $D_F(\boldsymbol{\pi}_{\mathbf{p}}, \mathbf{p}) = D_F(\boldsymbol{\pi}_{\mathbf{p}}, \mathbf{q})$ (definition of the bisector), we would have $D_F(\boldsymbol{\pi}_{\mathbf{q}}, \mathbf{q}) > D_F(\boldsymbol{\pi}_{\mathbf{p}}, \mathbf{q})$, contradicting the fact that $\boldsymbol{\pi}_{\mathbf{q}}$ is the projection of $\mathbf{q}$ on H. Finally, we say that $I \subseteq \mathcal{X}$ is *Bregman orthogonal* to $J \subseteq \mathcal{X}$ $(I \cap J \neq \emptyset)$ iff $\forall \mathbf{x} \in I, \boldsymbol{\pi} \in I \cap J, \mathbf{c} \in J$, $(\mathbf{c}, \boldsymbol{\pi}, \mathbf{x})$ satisfies the generalized Pythagorean theorem, that can be rewritten as $\langle \mathbf{c} - \boldsymbol{\pi}_{\mathbf{x}}, \boldsymbol{\pi}'_{\mathbf{x}} - \mathbf{x}' \rangle = 0$.

LEMMA 3.3. $\mathrm{c}(\mathbf{p}, \mathbf{q})$ *is Bregman orthogonal to* H. H^* *is Bregman orthogonal to* $\mathrm{l}(\mathbf{p}, \mathbf{q})$.

Proof. Remark first that $\mathrm{c}(\mathbf{p}, \mathbf{q}) \cap H \neq \emptyset$. Indeed, consider for example the Bregman orthogonal projection $\boldsymbol{\pi}_{\mathbf{p}}$ of $\mathbf{p}$ onto H. It may be retrieved via the Lagrangian $L(\mathbf{c}, \lambda) = D_F(\mathbf{c}, \mathbf{p}) + \lambda(\langle \mathbf{c}, \mathbf{p}' - \mathbf{q}' \rangle + F(\mathbf{p}) - \langle \mathbf{p}, \mathbf{p}' \rangle - F(\mathbf{q}) + \langle \mathbf{q}, \mathbf{q}' \rangle)$, with λ a Lagrange multiplier. Solving $\nabla_{\mathbf{c}} L(\mathbf{c}, \lambda)|_{\mathbf{c}=\boldsymbol{\pi}_{\mathbf{p}}} = 0$ yields $\boldsymbol{\pi}'_{\mathbf{p}} - \mathbf{p}' + \lambda(\mathbf{p}' - \mathbf{q}') = 0$, *i.e.* $\boldsymbol{\pi}'_{\mathbf{p}} = (1-\lambda)\mathbf{p}' + \lambda\mathbf{q}' \in \mathrm{c}(\mathbf{p}, \mathbf{q}) \cap H$. Now, fix any distinct $\mathbf{x} \in \mathrm{c}(\mathbf{p}, \mathbf{q})$, $\boldsymbol{\pi} \in \mathrm{c}(\mathbf{p}, \mathbf{q}) \cap H$ and $\mathbf{c} \in H$. It follows $\boldsymbol{\pi}' - \mathbf{x}' = \beta(\mathbf{p}' - \mathbf{q}')$ $(\beta \in \mathbb{R}^*)$, and, using the equation of H, we obtain $\langle \mathbf{c} - \boldsymbol{\pi}, \boldsymbol{\pi}' - \mathbf{x}' \rangle = \beta(\langle \mathbf{c}, \mathbf{p}' - \mathbf{q}' \rangle - \langle \pi, \mathbf{p}' - \mathbf{q}' \rangle) = \beta(-F(\mathbf{p}) + \langle \mathbf{p}, \mathbf{p}' \rangle + F(\mathbf{q}) - \langle \mathbf{q}, \mathbf{q}' \rangle + F(\mathbf{p}) - \langle \mathbf{p}, \mathbf{p}' \rangle - F(\mathbf{q}) + \langle \mathbf{q}, \mathbf{q}' \rangle) = 0$, *i.e.* $(\mathbf{c}, \boldsymbol{\pi}, \mathbf{x})$ satisfies the generalized Pythagorean theorem.

Now, consider any $\mathbf{x} \in H^*$ and $\boldsymbol{\pi}_{\mathbf{x}}$, its Bregman orthogonal projection on $\mathrm{l}(\mathbf{p}, \mathbf{q})$. We have $\langle \mathbf{p} - \boldsymbol{\pi}_{\mathbf{x}}, \boldsymbol{\pi}'_{\mathbf{x}} - \mathbf{x}' \rangle = 0$ and $\langle \mathbf{q} - \boldsymbol{\pi}_{\mathbf{x}}, \boldsymbol{\pi}'_{\mathbf{x}} - \mathbf{x}' \rangle = 0$, which yields $\langle \mathbf{p} - \mathbf{q}, \boldsymbol{\pi}'_{\mathbf{x}} \rangle = \langle \mathbf{p} - \mathbf{q}, \mathbf{x}' \rangle$. Using the equation of H^* (Lemma 3.1), it follows that $\boldsymbol{\pi}_{\mathbf{x}} \in H^*$, and $\mathrm{l}(\mathbf{p}, \mathbf{q}) \cap H^* \neq \emptyset$. Now, fix any distinct $\mathbf{c} \in \mathrm{l}(\mathbf{p}, \mathbf{q})$, $\boldsymbol{\pi} \in \mathrm{l}(\mathbf{p}, \mathbf{q}) \cap H^*$ and $\mathbf{x} \in H^*$. It follows $\mathbf{c} - \boldsymbol{\pi} = \beta(\mathbf{p} - \mathbf{q})$ $(\beta \in \mathbb{R}^*)$. Using the equation of H^*, we obtain $\langle \mathbf{c} - \boldsymbol{\pi}, \boldsymbol{\pi}' - \mathbf{x}' \rangle = \beta(\langle \boldsymbol{\pi}', \mathbf{p} - \mathbf{q} \rangle - \langle \mathbf{x}', \mathbf{p} - \mathbf{q} \rangle) = \beta(F(\mathbf{p}) - F(\mathbf{q}) - F(\mathbf{p}) + F(\mathbf{q})) = 0$, *i.e.* $(\mathbf{c}, \boldsymbol{\pi}, \mathbf{x})$ satisfies the generalized Pythagorean theorem.

Lemma 3.1 shows that the definitions of the first and second-type Bregman bisectors permute each others by $\boldsymbol{\nabla}_F$, and so their constructions are *dual* to each

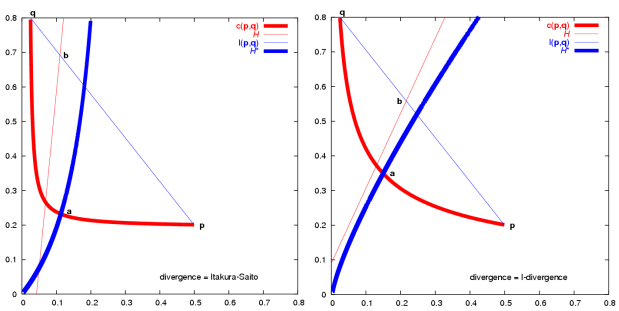

Figure 4: Bregman bisectors and their relationships with respect to $\mathrm{l}(\mathbf{p},\mathbf{q})$ and $\mathrm{c}(\mathbf{p},\mathbf{q})$, for the I-divergence (left) and Itakura-Saito divergence (right, $d = 2$). Bold curves become linear in $\boldsymbol{\nabla}_F\mathcal{X}$; colors depict the orthogonality relationships of Lemma 3.3 (see text for $\mathbf{a}$ and $\mathbf{b}$).

other. This is the same situation for $\mathrm{c}(\mathbf{p},\mathbf{q})$ and $\mathrm{l}(\mathbf{p},\mathbf{q})$. Figure 4 displays an example, in which $\mathbf{a}$ and $\mathbf{b}$ are defined by $H^* \cap \mathrm{c}(\mathbf{p},\mathbf{q}) = \{\mathbf{a}\}$ and $H \cap \mathrm{l}(\mathbf{p},\mathbf{q}) = \{\mathbf{b}\}$, *i.e.* $\mathbf{a}' = (1-\alpha)\mathbf{p}' + \alpha\mathbf{q}'$ and $\mathbf{b} = \alpha\mathbf{p} + (1-\alpha)\mathbf{q}$, with $\alpha = D_F(\mathbf{q},\mathbf{p})/(D_F(\mathbf{p},\mathbf{q}) + D_F(\mathbf{q},\mathbf{p}))$. Since linear in the gradient space, $\mathrm{c}(.,.)$ defines the geodesics of $\boldsymbol{\nabla}_F\mathcal{X}$. As such, it also defines the geodesics of the first-type Voronoi diagram, those curves that allow one to define dynamically Voronoi diagrams by time-dependent growth regions from the generators. This is formally stated in the following Lemma.

LEMMA 3.4. *The geodesic between any pair of points $(\mathbf{p},\mathbf{q}) \in \mathcal{X} \times \mathcal{X}$ is the curve defined by $\{\mathbf{x} \in \mathcal{X} : \mathbf{x} = {\boldsymbol{\nabla}_F}^{-1}((1-\lambda)\mathbf{p}' + \lambda\mathbf{q}'), \lambda \in [0,1]\}$.*

Naturally, $\mathrm{l}(.,.)$ defines the geodesics of second-type Voronoi diagram. Geodesics for the first-type Voronoi diagram turn out to describe the common weighted averages of $\mathbf{p}$ and $\mathbf{q}$ used in many domains, such as the arithmetic average for L_2^2 as the Bregman divergence, the geometric average for the I-divergence, and the harmonic average for Itakura-Saito divergence (see [19]).

3.2 Symmetrized Bregman Voronoi diagrams

We can further define the Voronoi diagram of the *symmetrized divergence* S_F (third-type Bregman Voronoi diagram) associated to the Bregman divergence D_F by choosing the pseudo-distance function S_F as $S_F(\mathbf{p},\mathbf{q}) = S_F(\mathbf{q},\mathbf{p}) = \frac{D_F(\mathbf{p},\mathbf{q})+D_F(\mathbf{q},\mathbf{p})}{2} = \frac{\langle \mathbf{p}-\mathbf{q}, \mathbf{p}'-\mathbf{q}' \rangle}{2}$. Divergence S_F can be handled as a Bregman divergence in dimension $2d$, where $d = \dim \mathcal{X}$. Indeed, let $\tilde{\mathbf{x}} = [\mathbf{x}\ \mathbf{x}']^T$ be the $2d$-dimensional vector obtained by stacking the coordinates of $\mathbf{x}$ on top of $\mathbf{x}'$, we have $S_F(\mathbf{p},\mathbf{q}) = \frac{D_F(\mathbf{p},\mathbf{q})+D_F(\mathbf{q},\mathbf{p})}{2} = \frac{D_F(\mathbf{p},\mathbf{q})+D_{F^*}(\mathbf{p}',\mathbf{q}')}{2} = D_{\tilde{F}}(\begin{bmatrix}\mathbf{p}\\ \mathbf{p}'\end{bmatrix}, \begin{bmatrix}\mathbf{q}\\ \mathbf{q}'\end{bmatrix})$, where $\tilde{F}(\tilde{\mathbf{x}}) = [F\ F^*]\begin{bmatrix}\mathbf{x}\\ \mathbf{x}'\end{bmatrix}$ for F^* the conjugate function of F ($D_{\tilde{F}} : \mathcal{X} \times \boldsymbol{\nabla}_F\mathcal{X} \mapsto \mathbb{R}$). We have $S_F(\mathbf{p},\mathbf{q}) = D_{\tilde{F}}(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$. That is, the symmetrized Bregman divergence of d-variate function F is a Bregman divergence $\tilde{F}$ of dimension $2d$. It follows that the third-type symmetrized Voronoi diagram with Voronoi cells $\mathrm{Vor}_F(\mathbf{p}_i) \stackrel{\mathrm{def}}{=} \{\mathbf{x} \in \mathcal{X} \mid S_F(\mathbf{x},\mathbf{p}_i) \leq S_F(\mathbf{x},\mathbf{p}_j)\ \forall \mathbf{p}_j \in \mathcal{S}\}$ is a Bregman Voronoi diagram: $\mathrm{Vor}_F(\mathcal{S}) \equiv \mathrm{vor}_{\tilde{F}}(\mathcal{S}) \equiv \mathrm{vor}_{\tilde{F}^*}(\mathcal{S}) \equiv \mathrm{vor}'_{\tilde{F}^*}(\mathcal{S}) \equiv \mathrm{vor}'_{\tilde{F}}(\mathcal{S})$.

4 Bregman Voronoi diagrams from polytopes

In this section, we consider Bregman Voronoi diagrams of the first-type, except when explicitly mentioned.

4.1 The lifting map

Let us embed the domain $\mathcal{X}$ in $\mathcal{X}^+ = \mathcal{X} \times \mathbb{R}^+$ using an extra dimension denoted by the Z-axis. The graph of F is the convex hypersurface of $\mathcal{X}^+$: $\mathcal{F} : z = F(\mathbf{x})$. For a point $\mathbf{x} \in \mathcal{X}$, let $\hat{\mathbf{x}}$ denote its corresponding point in $\mathcal{F} \subset \mathcal{X}^+$ obtained by the lifting transformation $\hat{\mathbf{x}} = (\mathbf{x}, F(\mathbf{x}))$. In addition, we write $\mathrm{Proj}_\perp(\mathbf{x},z) = \mathbf{x}$ for the vertical projection (orthographic projection) of a point $(\mathbf{x},z) \in \mathcal{X}^+$ obtained by dropping the extra vector z-coordinate.

Let $\mathbf{p} \in \mathcal{X}$ and $H_\mathbf{p}$ be the hyperplane tangent to $\mathcal{F}$ at point $\hat{\mathbf{p}}$ of equation $z = H_\mathbf{p}(\mathbf{x}) = \langle \mathbf{x}-\mathbf{p}, \mathbf{p}' \rangle + F(\mathbf{p})$, and let $H_\mathbf{p}^\uparrow$ denote the halfspace above $H_\mathbf{p}$ consisting of the points $\mathbf{x} = [\mathbf{x}\ z]^T \in \mathcal{X}^+$ such that $z > H_\mathbf{p}(\mathbf{x})$. A simple but key observation is the following (see Fig. 1)

LEMMA 4.1. *$D_F(\mathbf{x},\mathbf{y})$ is measured by the vertical distance from $\hat{\mathbf{x}}$ to $H_\mathbf{y}$.*

Proof. $D_F(\mathbf{x},\mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \langle \mathbf{x}-\mathbf{y}, \mathbf{y}' \rangle = F(\mathbf{x}) - H_\mathbf{y}(\mathbf{x})$.

4.2 Bregman spheres and polarity

We define the (first-type) Bregman sphere of $\mathcal{X}$ of center $\mathbf{c}$ and radius r as $\sigma(\mathbf{c},r) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x},\mathbf{c}) = r\}$ and the associated ball $B(\mathbf{c},r) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x},\mathbf{c}) \leq r\}$. The lifted image $\hat{\sigma}$ of a Bregman sphere σ is $\hat{\sigma} = \{(\mathbf{x}, F(\mathbf{x})), \mathbf{x} \in \sigma\}$. We associate to a Bregman sphere $\sigma = \sigma(\mathbf{c},r)$ of $\mathcal{X}$ the hyperplane $H_\sigma : z = \langle \mathbf{x}-\mathbf{c}, \mathbf{c}' \rangle + F(\mathbf{c}) + r$, parallel to $H_\mathbf{c}$ and at vertical distance r from $H_\mathbf{c}$. Observe that H_σ coincides with $H_\mathbf{c}$ when $r = 0$, i.e. when σ is a point.

LEMMA 4.2. *$\hat{\sigma}$ is the intersection of $\mathcal{F}$ with H_σ. Conversely, the intersection of any hyperplane H with $\mathcal{F}$ projects onto $\mathcal{X}$ as a Bregman sphere. More precisely, if the equation of H is $z = \langle \mathbf{x}, \mathbf{a} \rangle + b$, the sphere is centered at ${\boldsymbol{\nabla}_F}^{-1}(\mathbf{a})$ and its radius is $\langle {\boldsymbol{\nabla}_F}^{-1}(\mathbf{a}), \mathbf{a} \rangle - F({\boldsymbol{\nabla}_F}^{-1}(\mathbf{a})) + b$.*

Proof. The first part of the lemma is a direct consequence of Lemma 4.1. For the second part, we consider the hyperplane H' parallel to H and tangent to $\mathcal{F}$. The point where H is tangent to $\mathcal{F}$ is the point ${\nabla_F}^{-1}(\mathbf{a})$. Hence, the equation of H' is $z = \langle \mathbf{x} - {\nabla_F}^{-1}(\mathbf{a}), \mathbf{a} \rangle + F({\nabla_F}^{-1}(\mathbf{a}))$. Using Lemma 4.1 again allows to conclude.

Let us restrict our attention to symmetric Bregman divergences, i.e. $D_F(\mathbf{x}, \mathbf{y}) = D_F(\mathbf{y}, \mathbf{x})$ for any $\mathbf{x}$ and $\mathbf{y}$. It then follows from the two previous lemmata that all the hyperplanes tangent to $\mathcal{F}$ at the points of $\hat{\sigma}$ intersect in a common point, namely $\sigma^+ = (\mathbf{c}, F(c) - r)$. We call σ^+ the *pole* of H_σ and H_σ the *polar hyperplane* of σ^+. When $F(\mathbf{x}) = ||\mathbf{x}||^2$, $\mathcal{F}$ is a *paraboloid of revolution* and the correspondence between σ^+ and H_σ is the usual polarity with respect to the paraboloid [9].

LEMMA 4.3. *Polarity preserves incidences, i.e., for any two spheres σ_1 and σ_2, $\sigma_1^+ \in H_{\sigma_2} \Leftrightarrow {\sigma_2}^+ \in H_{\sigma_1}$.*

Proof.

$$\begin{aligned} \sigma_1^+ \in H_{\sigma_2} &\Leftrightarrow F(\mathbf{c}_1) - r_1 = \langle \mathbf{c}_1 - \mathbf{c}_2, \mathbf{c}_2' \rangle + F(\mathbf{c}_2) + r_2 \\ &\Leftrightarrow D_F(\mathbf{c}_1, \mathbf{c}_2) = r_1 + r_2 = D_F(\mathbf{c}_2, \mathbf{c}_1) \\ &\Leftrightarrow \sigma_2^+ \in H_{\sigma_1} \end{aligned}$$

We mention another important property of Bregman spheres in the following lemma:

LEMMA 4.4. (PROOF IN [18]) *Let $\sigma(\mathbf{c}, r)$ be a Bregman sphere. Any geodesic passing through the center $\mathbf{c}$ is Bregman orthogonal to σ at exactly two antipodal points.*

4.3 Bregman Voronoi diagrams from polytopes Let $\mathbf{p}_1$ and $\mathbf{p}_2$ be two points. According to Lemma 4.1, the intersection of the two hyperplanes $H_{\mathbf{p}_1}$ and $H_{\mathbf{p}_2}$ projects onto $\mathcal{X}$ along the Bregman bisector of $\mathbf{p}_1$ and $\mathbf{p}_2$. Consider $d+1$ points $\mathbf{p}_0, \ldots, \mathbf{p}_d$. The affine hull of these points is a hyperplane of $\mathcal{X}^+$ whose vertical projection coincides with the (unique) Bregman sphere passing through $\mathbf{p}_0, \ldots, \mathbf{p}_d$. The corresponding hyperplanes $H_{\mathbf{p}_0}, \ldots, H_{\mathbf{p}_d}$ intersect in a single point whose projection is the (unique) Bregman circumcenter of the points. Consider now a set $\mathcal{S}$ of n points $\mathbf{p}_1, \ldots, \mathbf{p}_n$ and another point $\mathbf{x}$. According to Lemma 4.1, the point of $\mathcal{S}$ which is closest to $\mathbf{x}$ (i.e. that minimizes $D_F(\mathbf{x}, .)$) is the point $\mathbf{p}_i$ such that $H_{\mathbf{p}_i}$ intersects the vertical line passing through $\mathbf{p}$ above all the other hyperplanes $H_{\mathbf{p}_j}$, $j \neq i$. The following theorem follows.

THEOREM 4.1. *The first-type Bregman Voronoi diagram $\mathrm{vor}_F(\mathcal{S})$ is obtained by projecting by $\mathrm{Proj}_\perp$ the faces of the $(d+1)$-dimensional polytope $\mathcal{H} = \cap_i H^\uparrow_{\mathbf{p}_i}$ of $\mathcal{X}^+$ onto $\mathcal{X}$.*

From McMullen's upperbound theorem [17] and Chazelle's optimal algorithm [10], we know that a polytope of $\mathbb{R}^d$ defined as the intersection of n halfspaces has complexity $\Theta(n^{\lfloor \frac{d}{2} \rfloor})$ and can be computed in optimal-time $\Theta(n \log n + n^{\lfloor \frac{d}{2} \rfloor})$ for any fixed dimension d. From Theorem 4.1 and Lemma 3.2, we then deduce the following theorem.

THEOREM 4.2. *The Bregman Voronoi diagrams of type 1 or 2 of a set of n d-dimensional points have complexity $\Theta(n^{\lfloor \frac{d+1}{2} \rfloor})$ and can be computed in optimal time $\Theta(n \log n + n^{\lfloor \frac{d+1}{2} \rfloor})$. The third-type Bregman Voronoi diagram for the symmetrized Bregman divergence of a set of n d-dimensional points has complexity $\Theta(n^d)$ and can be computed in optimal time $\Theta(n^d)$ (for $d \geq 2$).*

4.4 Bregman triangulations Let $\hat{\mathcal{S}}$ be the lifted image of $\mathcal{S}$ and let $\mathcal{T}$ be the lower convex hull of $\hat{\mathcal{S}}$, i.e. the collection of facets of the convex hull of $\hat{\mathcal{S}}$ whose supporting hyperplanes are below $\hat{\mathcal{S}}$. We say that $\mathcal{S}$ is in general position if there is no subset of $d+2$ points lying on a same Bregman sphere. Equivalently (see Lemma 4.2), $\mathcal{S}$ is in general position if no subset of $d+2$ hyperplanes $H_{\mathbf{p}_i}$ intersect in a single point.

In the rest of the paper, we assume that $\mathcal{S}$ is in *general position*. Then, each vertex of $\mathcal{H}$ is the intersection of exactly $d+1$ hyperplanes and the faces of $\mathcal{T}$ are all simplices. Moreover the vertical projection of $\mathcal{T}$ is a triangulation $BT(\mathcal{S}) = \mathrm{Proj}_\perp(\mathcal{T})$ of $\mathcal{S}$ embedded in $\mathbb{R}^d$. Indeed, since the restriction of $\mathrm{Proj}_\perp$ to $\mathcal{T}$ is bijective, $BT(\mathcal{S})$ is a simplicial complex embedded in $\mathcal{X}$. Moreover, since F is convex, $BT(\mathcal{S})$ covers the convex hull of $\mathcal{S}$ and the set of vertices of $\mathcal{T}$ consists of all the $\hat{\mathbf{p}}_i$. Consequently, the set of vertices of $BT(\mathcal{S})$ is $\mathcal{S}$. We call $BT(\mathcal{S})$ the *Bregman triangulation* of $\mathcal{S}$ (see Fig. 5). When $F(\mathbf{x}) = ||\mathbf{x}||^2$, $BT(\mathcal{S})$ is the well-known Delaunay triangulation dual to the Euclidean Voronoi diagram. We have a similar result for symmetric Bregman divergences.

LEMMA 4.5. *For symmetric Bregman divergences, the Bregman triangulation of $\mathcal{S}$ is dual to the Bregman Voronoi diagram of $\mathcal{S}$.*

Proof. For symmetric Bregman divergences, the polarity introduced in Section 4.2 provides a bijective mapping between the faces of $\mathcal{H}$ and the faces of $\mathcal{T}$. More precisely, the mapping associates to a face $f = \cap_{i \in \nu} H_{\mathbf{p}_i}$ of $\mathcal{H}$ the dual face f^* that is the convex hull of the $\hat{\mathbf{p}}_i$, $i \in \nu$. This duality is an involution, i.e. $f^{**} = f$, and satisfies $f \subset g \Rightarrow g^* \subset f^*$. Hence, $BT(\mathcal{S})$ is dual to the Bregman Voronoi diagram of $\mathcal{S}$.

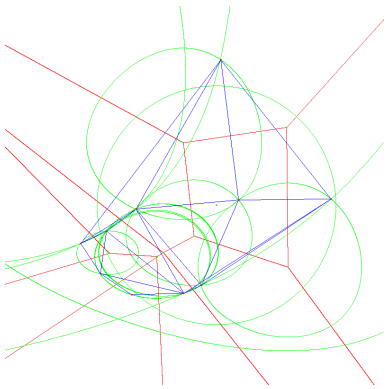

Figure 5: A 2D Bregman Voronoi diagram and its dual regular triangulation for the exponential divergence.

We say that a Bregman sphere σ is *empty* if the open ball bounded by σ does not contain any point of $\mathcal{S}$. The following theorem extends a similar well-known property for Delaunay triangulations whose proof (see, for example, [7]) can be extended in a straightforward way to Bregman triangulations.

THEOREM 4.3. *The Bregman sphere circumscribing any simplex of $BT(\mathcal{S})$ is empty. $BT(\mathcal{S})$ is the only triangulation of $\mathcal{S}$ with this property when $\mathcal{S}$ is in general position.*

Several other properties of Delaunay triangulations extend to Bregman triangulations. We list some of them.

THEOREM 4.4. (EMPTY BALL) *Let ν be a subset of at most $d+1$ indices in $\{1,\ldots,n\}$. The convex hull of the associated points $\mathbf{p}_i$, $i \in \nu$, is a simplex of the Bregman triangulation of $\mathcal{S}$ iff there exists an empty Bregman sphere σ passing through the $\mathbf{p}_i$, $i \in \nu$.*

The next property exhibits a local characterization of Bregman triangulations. Let $T(\mathcal{S})$ be a triangulation of $\mathcal{S}$. We say that a pair of adjacent facets $f_1 = (f, \mathbf{p}_1)$ and $f_2 = (f, \mathbf{p}_2)$ of $T(\mathcal{S})$ is regular iff $\mathbf{p}_1$ does not belong to the open Bregman ball circumscribing f_2 and $\mathbf{p}_2$ does not belong to the open Bregman ball circumscribing f_1 (the two statements are equivalent for symmetric Bregman divergences).

THEOREM 4.5. (LOCALITY) *Any triangulation of a given set of points $\mathcal{S}$ (in general position) whose pairs of facets are all regular is the Bregman triangulation of $\mathcal{S}$.*

Let $\mathcal{S}$ be a given set of points, $BT(\mathcal{S})$ its Bregman triangulation, and $\mathcal{T}(\mathcal{S})$ the set of all the triangulations of $\mathcal{S}$. We define the Bregman radius of a d-simplex τ as the radius noted $r(\tau)$ of the smallest Bregman ball containing τ. The following result is an extension of a result due to Rajan for Delaunay triangulations [22].

THEOREM 4.6. (OPTIMALITY) *We have* $BT(\mathcal{S}) = \min_{T \in \mathcal{T}(\mathcal{S})} \max_{\tau \in T} r(\tau)$.

The proof mimics Rajan's proof [22] for the case of Delaunay triangulations.

5 Bregman Voronoi diagrams from Power diagrams

5.1 Power diagrams and regular triangulations

The power distance of a point $\mathbf{x}$ to a ball $B = \text{Ball}(\mathbf{p}, r)$ is defined as $||\mathbf{p} - \mathbf{x}||^2 - r^2$. Given n balls $B_i = \text{Ball}(\mathbf{p}_i, r_i)$, $i = 1,\ldots,n$, the power diagram of n balls is defined as the minimization diagram of the corresponding n functions $||\mathbf{p}_i - \mathbf{x}||^2 - r^2$. The power bisector of any two balls $\text{Ball}(\mathbf{p}, r_p)$ and $\text{Ball}(\mathbf{q}, r_q)$ is a hyperplane of equation $2\langle \mathbf{x}, \mathbf{q} - \mathbf{p}\rangle + ||\mathbf{p}||^2 - ||\mathbf{q}||^2 + r_q^2 - r_p^2 = 0$. Thus power diagrams are affine diagrams. Moreover, as shown by Aurenhammer [2, 7], any affine diagram is identical to the power diagram of a set of balls.

It is well known that a power diagram of a finite set of balls of $\mathbb{R}^d$ has a dual triangulation known as the *regular triangulation*. This triangulation is embedded in $\mathbb{R}^d$ with its vertices at the centers of the hyperspheres.

In general, some balls may have an empty cell in their power diagram. Equivalently, some sites may not appear as vertices of the dual regular triangulation.

5.2 Bregman Voronoi diagrams from power diagrams

Since Bregman Voronoi diagrams of the first type are affine diagrams, Bregman Voronoi diagrams are power diagrams [2, 7]. Observe however that Bregman Voronoi diagrams are special cases of power diagrams since, differently from power diagrams, any cell in a Bregman Voronoi diagram contains its generator and therefore is non empty. See Section 6.1 for a further discussion on this point. The following theorem makes precise this correspondence.

THEOREM 5.1. *The first-type Bregman Voronoi diagram of n sites of $\mathcal{X}$ is identical to the power diagram of the n Euclidean hyperspheres of equations $\langle \mathbf{x} - \mathbf{p}'_i, \mathbf{x} - \mathbf{p}'_i\rangle = \langle \mathbf{p}'_i, \mathbf{p}'_i\rangle + 2(F(\mathbf{p}_i) - \langle \mathbf{p}_i, \mathbf{p}'_i\rangle)$ for $i = 1,\ldots,n$.*

Proof. Consider wlog. $D_F(\mathbf{x}, \mathbf{p}_i) \leq D_F(\mathbf{x}, \mathbf{p}_j)$. That is, $-F(\mathbf{p}_i) - \langle \mathbf{x} - \mathbf{p}_i, \mathbf{p}'_i\rangle \leq -F(\mathbf{p}_j) - \langle \mathbf{x} - \mathbf{p}_j, \mathbf{p}'_j\rangle$. It comes that $\langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{x}, \mathbf{p}'_i\rangle - 2F(\mathbf{p}_i) + 2\langle \mathbf{p}_i, \mathbf{p}'_i\rangle \leq \langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{x}, \mathbf{p}'_j\rangle - 2F(\mathbf{p}_j) + 2\langle \mathbf{p}_j, \mathbf{p}'_j\rangle$. Thus, $\langle \mathbf{x} - \mathbf{p}'_i, \mathbf{x} - \mathbf{p}'_i\rangle -$

$r_i^2 \leq \langle \mathbf{x}-\mathbf{p}'_j, \mathbf{x}-\mathbf{p}'_j\rangle - r_j^2$, where $r_i^2 = \langle \mathbf{p}'_i, \mathbf{p}'_i\rangle + 2(F(\mathbf{p}_i) - \langle \mathbf{p}_i, \mathbf{p}'_i\rangle)$ and $r_j^2 = \langle \mathbf{p}'_j, \mathbf{p}'_j\rangle + 2(F(\mathbf{p}_j) - \langle \mathbf{p}_j, \mathbf{p}'_j\rangle)$. The last inequality means that the power of $\mathbf{x}$ with respect to the Euclidean (possibly imaginary) sphere $S(\mathbf{p}'_i, r_i)$ is no more than the power of $\mathbf{x}$ with respect to the Euclidean (possibly imaginary) sphere $S(\mathbf{p}'_j, r_j)$.

Note that for $F(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, \mathbf{x}\rangle$ (half the squared Euclidean distance D_F) that yields the ordinary Voronoi diagram as well, we have $\langle \mathbf{x}, \mathbf{x}'\rangle - F(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, \mathbf{x}\rangle$ and $\mathbf{x}' = \mathbf{x}$, and therefore get $r_i^2 = r_j^2 = 0$ as expected.

5.3 Bregman geodesic triangulations. The regular triangulation dual to the power diagram above is a triangulation of the points $\mathbf{p}'_i$. The image of this triangulation by ∇_F^{-1} is a *curved triangulation* whose vertices are the $\mathbf{p}_i$. The edges of this curved triangulation are geodesic arcs joining two sites by Lemma 3.4. The reduction to power diagrams provides an alternative convenient solution for computing Bregman Voronoi diagrams as many available software libraries handle power diagrams (*e.g.*, QHULL, CGAL or LEDA). Note however that computing the reciprocal gradient for computing the curved Bregman Voronoi diagrams is a non-trivial operation that can be quite challenging for some convex functions F.

6 Generalized Bregman divergences and their Voronoi diagrams

6.1 Weighted Bregman Voronoi diagrams We associate to each site $\mathbf{p}_i$ a weight $w_i \in \mathbb{R}$. We define the weighted divergence between two weighted points as $WD_F(\mathbf{p}_i, \mathbf{p}_j) \stackrel{\text{def}}{=} D_F(\mathbf{p}_i, \mathbf{p}_j) + w_i - w_j$. We can define bisectors and weighted Bregman Voronoi diagram in very much the same way as for non weighted divergences: $\text{vor}_F(\mathbf{p}_i, w_i) \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x}, \mathbf{p}_i) + w_i \leq D_F(\mathbf{x}, \mathbf{p}_j) + w_j \ \forall \mathbf{p}_j \in \mathcal{S}\}$. Observe that the bisectors of the 1st-type diagrams are still hyperplanes and that the diagram can be obtained as the projection of a H-polytope or as the power diagram of a finite set of balls. The only difference wrt the construction of Section 4 is the fact that now the hyperplanes $H_{\mathbf{p}_i}$ are no longer tangent to $\mathcal{F}$: they are indeed *shifted* by a z-displacement of length w_i. As a consequence, the cells of some sites may be empty and the class of weighted Bregman Voronoi diagrams is identical to the class of affine or power diagrams.

THEOREM 6.1. *The weighted Bregman Voronoi diagrams of type 1 or 2 of a set of n d-dimensional points have complexity $\Theta(n^{\lfloor \frac{d+1}{2} \rfloor})$ and can be computed in optimal time $\Theta(n \log n + n^{\lfloor \frac{d+1}{2} \rfloor})$.*

We define the k-order Bregman Voronoi diagram of n punctual sites of $\mathcal{X}$ as the subdivision of $\mathcal{X}$ into cells such that each cell is associated to a subset of k sites and consists of the points of $\mathcal{X}$ whose divergence to any site in the subset is less than the divergence to the other sites. Similarly to the case of higher-order Euclidean Voronoi diagrams, we have:

THEOREM 6.2. *The k-order Bregman Voronoi diagram of n d-dimensional points is a weighted Bregman Voronoi diagram.*

Proof. Let $P_1, P_2, \ldots$ denote the subsets of k points of P and write $D_i(\mathbf{x}) = \frac{1}{k}\sum_{j\in P_i} D_F(\mathbf{x}, \mathbf{p}_j)$. That is $D_i(\mathbf{x}) = F(\mathbf{x}) - \frac{1}{k}\sum_{j\in P_i} F(\mathbf{p}_j) + \frac{1}{k}\sum_{j\in P_i}\langle \mathbf{x}-\mathbf{p}_j, \mathbf{p}'_j\rangle$. Thus, $D_i(\mathbf{x}) = F(\mathbf{x}) - F(\mathbf{c}_i) - \langle \mathbf{x}-\mathbf{c}_i, \mathbf{c}'_i\rangle + w_i = WD_F(\mathbf{x}, \mathbf{c}_i)$, where $\mathbf{c}_i = {\nabla_F}^{-1}\left(\frac{1}{k}\sum_{j\in P_i}\mathbf{p}'_j\right)$ and the weight associated to $\mathbf{c}_i$ is $w_i = F(\mathbf{c}_i) - \langle \mathbf{c}_i, \mathbf{c}'_i\rangle - \frac{1}{k}\sum_{j\in P_i}\left(F(\mathbf{p}_j) + \langle \mathbf{p}_j, \mathbf{p}'_j\rangle\right)$.

Hence, P_i is the set of the k nearest neighbors of $\mathbf{x}$ iff $\forall j,\quad D_i(\mathbf{x}) \leq D_j(\mathbf{x})$, or equivalently, iff $\mathbf{x}$ belongs to the cell of $\mathbf{c}_i$ in the weighted Bregman Voronoi diagram of the $\mathbf{c}_i$.

6.2 Higher-jet divergence Let us now turn on to the notion of higher-jet divergences. Let $D_F^{(k)}$ be a *k-jet divergence* obtained from the tail of a Taylor expansion of a convex function F: $D_F^{(k)}(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \sum_{i=1}^{k} \frac{D^i F(\mathbf{y})(\mathbf{x}-\mathbf{y})^i}{i!}$ (multi-index notation with $\sum_{i=1}^{k} \frac{D^i F(\mathbf{y})(\mathbf{x}-\mathbf{y})^i}{i!}$ the truncated Taylor expansion). Let us write for short $D_F = D_F^{(1)}$ the standard Bregman divergence (i.e. $k = 1$) and $D_F^{(2)}(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - (\mathbf{x}-\mathbf{y})^T\mathbf{y}' - (\mathbf{x}-\mathbf{y})^T\frac{{\nabla_F}^2(\mathbf{y})}{2}(\mathbf{x}-\mathbf{y}) = D_F^{(1)}(\mathbf{x}, \mathbf{y}) - (\mathbf{x}-\mathbf{y})^T\mathbf{Q}_F(\mathbf{y})\,(\mathbf{x}-\mathbf{y})$ where $\mathbf{Q}_F(\mathbf{y}) = \frac{1}{2}{\nabla_F}^2(\mathbf{y})$ denote half of the Hessian symmetric matrix of F defined over $\mathbb{R}^d$.

The second-jet divergence not being symmetric, we define the bisector of two points $\mathbf{p}$ and $\mathbf{q}$ in two different ways $H(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid Q_F(\mathbf{x}, \mathbf{p}) = Q_F(\mathbf{x}, \mathbf{q})\}$ (first-type) and $H'(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid Q_F(\mathbf{p}, \mathbf{x}) = Q_F(\mathbf{q}, \mathbf{x})\}$ (second-type). Plainly, a bisector of the first-type is a quadratic hypersurface of $\mathbb{R}^d$ and the associated second-jet Voronoi diagram of the first type is therefore an *anisotropic* Voronoi diagram [6].

Let q_{ij}, $i = 1, \cdots, d, j = 1, \cdots, d$ be the elements of $\mathbf{Q}_F(\mathbf{y})$ and $\bar{\mathbf{q}}_F^t(\mathbf{y}) = (q_{ij}, i = 1, \cdots, d, i \leq j)$ the vector of the elements of the upper triangle of $\mathbf{Q}_F(\mathbf{y})$. Let $\mathbf{x}^t = (x_1, \cdots, x_d)$, $\bar{\mathbf{x}}^t = (\frac{1}{2^{\delta_{ij}}}x_i x_j, i = 1, \cdots, d, i \leq j)$ and $\mathbf{X}^t = (\mathbf{x}, \bar{\mathbf{x}})^t$ where δ_{ij} is 1 if $i = j$ and 0 otherwise [6]. Observe that $\mathbf{X} \in \mathbb{R}^D$ where $D = \frac{d(d+3)}{2}$. We have $(\mathbf{x}-\mathbf{y})^t\,\mathbf{Q}_F(\mathbf{y})\,(\mathbf{x}-\mathbf{y}) = 2(\bar{\mathbf{x}}-\bar{\mathbf{y}})^t\bar{\mathbf{q}}_F(\mathbf{y})$ and therefore $D_F^{(2)}(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - (\mathbf{x}-\mathbf{y})^t\mathbf{y}' - 2(\bar{\mathbf{x}}-\bar{\mathbf{y}})^t\bar{\mathbf{q}}_F(\mathbf{y}) =$

$F(\mathbf{x}) - F(\mathbf{y}) - (\mathbf{X} - \mathbf{Y})^t \begin{pmatrix} \mathbf{y}' \\ 2\bar{\mathbf{q}}_F(\mathbf{y}) \end{pmatrix} = PD_F(\mathbf{X}, \mathbf{Y})$. The expression of $PD_F(\mathbf{X}, \mathbf{Y})$ is similar to (though different from) the definition of a Bregman divergence. In particular, first-type bisectors are linear in $\mathbf{X}$ and therefore hyperplanes of $\mathbb{R}^D$. Let ϕ be the mapping: $\mathbf{x} \in \mathbb{R}^d \mapsto \mathbf{X} \in \mathbb{R}^D$. The above discussion together with section 3 leads to the following theorem (see also [6]).

THEOREM 6.3. *The first-type second-jet Voronoi diagram of n sites of $\mathbb{R}^d$ can be obtained as the pull-back by ϕ of a power diagram of n spheres in $\mathbb{R}^D$, $D = \frac{d(d+3)}{2}$. By the convex duality, the same result holds for the second-type second-jet Voronoi diagram.*

The combinatorial complexity of those diagrams is $O(n^{d+\varepsilon})$ by a result of Sharir [25] but smaller for some special cases of $\mathbf{Q}_F$. We can distinguish the *isotropic* case : $\mathbf{Q}_F(\mathbf{y}) = \sigma(\mathbf{y})\mathbf{I}$ where $\sigma(\mathbf{y}) \in \mathbb{R}$ and $\mathbf{I}$ is the $d \times d$ identity matrix. In this case, $D = d + 1$ and the second-jet Voronoi diagram is a Möbius diagram whose complexity is $\Theta(n^{\lfloor \frac{d+2}{2} \rfloor})$ [5]. When $\mathbf{Q}_F$ is a diagonal matrix, $D = 2d$ and the combinatorial complexity of the diagram is the same as the combinatorial complexity of a power diagram in $\mathbb{R}^{2d}$, specifically $\Theta(n^d)$. This bound is tight.

7 Applications

7.1 Union of Bregman balls

THEOREM 7.1. *The union of n Bregman balls of $\mathcal{X}$ has complexity $\Theta(n^{\lfloor \frac{d+1}{2} \rfloor})$ and can be computed in time $\Theta(n \log n + n^{\lfloor \frac{d+1}{2} \rfloor})$.*

Proof. To each ball, we can associate its bounding Bregman hypersphere σ_i which, by Lemma 4.2, is the projection by $\mathrm{Proj}_\perp$ of the intersection of $\mathcal{F}$ with a hyperplane H_{σ_i}. The points of $\mathcal{F}$ that are below H_{σ_i} projects onto points that are inside the ball bounded by σ_i. Hence the union of the balls is the projection by $\mathrm{Proj}_\perp$ of the complement of $\mathcal{F} \cap \left(\cap_{i=1}^n H_{\sigma_i}^\uparrow\right)$.

7.2 Centroidal Bregman Voronoi diagrams

We define the Bregman centroid of a domain $\mathcal{D} \subset \mathcal{X}$ as the point $\mathbf{c}^* \in \mathcal{X}$ such that $\mathbf{c}^* = \mathrm{argmin}_{\mathbf{c} \in \mathcal{D}} \int_{\mathbf{x} \in \mathcal{D}} D_F(\mathbf{x}, \mathbf{c})\, \mathrm{d}x$. The following lemma states that the Bregman centroid of $\mathcal{D}$ is uniquely defined and does not depend on F.

LEMMA 7.1. *The Bregman centroid of $\mathcal{D}$ coincides with the (L_2) centroid (center of mass) of $\mathcal{D}$.*

Proof. $\nabla \int_{\mathbf{x} \in \mathcal{D}} D_F(\mathbf{x}, \mathbf{c})\, \mathrm{d}x = \nabla \int_{\mathbf{x} \in \mathcal{D}} (F(\mathbf{x}) - F(\mathbf{c}) - \langle \mathbf{x} - \mathbf{c}, \boldsymbol{\nabla}_F(\mathbf{c}) \rangle) \mathrm{d}x = -\int_{\mathbf{x} \in \mathcal{D}} \boldsymbol{\nabla}_F{}^2(\mathbf{c})(\mathbf{x} - \mathbf{c}) \mathrm{d}x = -\boldsymbol{\nabla}_F{}^2(\mathbf{c})(\int_{\mathbf{x} \in \mathcal{D}} \mathbf{x} \mathrm{d}x - \mathbf{c} \int_{\mathbf{x} \in \mathcal{D}} \mathrm{d}x)$. Hence, $\mathbf{c}^* = \frac{\int_{\mathbf{x} \in \mathcal{D}} \mathbf{x} \mathrm{d}x}{\int_{\mathbf{x} \in \mathcal{D}} \mathrm{d}x}$.

Note that this result extends the discrete case (finite point sets) studied in [4]. Computing a centroidal Bregman Voronoi diagram of k points can be done by means of Lloyd's `kmeans` algorithm [16]. We select an initial set of k points. Then, we iteratively compute a Bregman Voronoi diagram and move the sites to the Bregman centroids of the corresponding cells in the diagram. The fixed point of this algorithm is a centroidal Bregman Voronoi diagram by standard arguments [11]. The output of the algorithm is a local minimizer of $f((\mathbf{p}_i, V_i), i = 1, \ldots, k) = \sum_{i=1}^k \int_{\mathbf{x} \in V_i} D_F(\mathbf{x}, \mathbf{p}_i)\, \mathrm{d}x$, where $\{\mathbf{p}_i\}_{i=1}^k$ denotes any set of k points of $\mathcal{X}$ and $\{V_i\}_{i=1}^k$ denotes any tesselation of $\mathcal{X}$ into k regions.

7.3 Sampling The `kmeans` algorithm [16] intends to find a best set of k points for a given k. Differently, we may want to sample a compact domain $\mathcal{D}$ up to a given precision while minimizing the number of samples. More precisely, let us define the error associated to a sample P as $\mathrm{error}(P) = \max_{\mathbf{x} \in \mathcal{D}} \min_{\mathbf{p}_i \in P} D_F(\mathbf{x}, \mathbf{p}_i)$. A finite set of points P of $\mathcal{D}$ is an ε-sample of $\mathcal{D}$ iff $\mathrm{error}(P) \leq \varepsilon$. For simplicity, we assume in the rest of the section that $\mathcal{D}$ is a convex polytope. Extending the results to more general domains is possible.

Let $P \subset \mathcal{D}$, $BVD(P)$ be the Bregman Voronoi diagram of P and $BVD_{\mathcal{D}}(P)$ be its restriction to $\mathcal{D}$. Write V for the set of vertices of the restriction of the 1-skeleton of $BVD_{\mathcal{D}}(P)$. V consists of vertices of $BVD(P)$ and intersection points between the edges of $BVD(P)$ and the boundary of $\mathcal{D}$. The following lemma states that $\mathrm{error}(P)$ can be computed by examining only a finite number of points, namely the points of V.

LEMMA 7.2. $\mathrm{error}(P) = \max_{\mathbf{v} \in V} \min_{\mathbf{p}_i \in P} D_F(\mathbf{x}, \mathbf{p}_i)$.

Proof. Let $\mathbf{x} \in \mathcal{D}$, $\mathbf{p}_\mathbf{x}$ the point of P closest to $\mathbf{x}$ and $V_\mathbf{x}$ the associated cell of $BVD_{\mathcal{D}}(P)$ (which contains $\mathbf{x}$). $V_\mathbf{x}$ is a bounded polytope whose vertices belongs to V. Let $\mathbf{w}$ be the vertex of $V_\mathbf{x}$ most distant from $\mathbf{p}_\mathbf{x}$. We have $D_F(\mathbf{x}, \mathbf{p}_\mathbf{x}) \leq D_F(\mathbf{w}, \mathbf{p}_\mathbf{x})$. This is a consequence of the convexity of F and of the fact that $D_F(\mathbf{x}, \mathbf{p})$ is measured by the vertical distance between $H_\mathbf{p}$ and $H_\mathbf{x}$ (Lemma 4.1).

The sampling problem is to find an ε-sample of minimal size. A simple solution to this problem is the following greedy algorithm originally proposed by Ruppert in the context of mesh generation [23]. See also [12]. We initialize the sample set P_0 with d points of $\mathcal{D}$ lying at distance greater than ε from one another. Then, at each step, the algorithm looks for the point $\mathbf{v}_i$ of $\mathcal{D}$ that is the furthest (for the considered Bregman divergence) from the current set of samples P_i. By Lemma 7.2, this step reduces to looking at the vertices of $BVD_{\mathcal{D}}(P_i)$. If

$D_F(\mathbf{x}, \mathbf{v}_i) \leq \varepsilon$, the algorithm stops. Otherwise, we take $\mathbf{v}_i$ as a new sample point, i.e. $\mathbf{p}_{i+1} = \mathbf{v}_i$, we update the set of sample points, i.e. $P_{i+1} = P_i \cup \{\mathbf{p}_{i+1}\}$, and insert $\mathbf{p}_{i+1}$ in the Bregman Voronoi diagram of the sample points. Upon termination, the set of sample points P_t satisfies the hypothesis of Lemma 7.2 and therefore P_t is an ε-sample of $\mathcal{D}$.

To prove that the algorithm terminates, we need the following lemma. Given a Bregman ball $B(\mathbf{c}, r)$, we define the biggest Euclidean ball $EB(\mathbf{c}, r')$ contained in $B(\mathbf{c}, r)$ and the smallest Euclidean ball $EB(\mathbf{c}, r'')$ containing $B(\mathbf{c}, r)$.

LEMMA 7.3. (PROOF IN [18]) *Let F be a strictly convex function of class C^2, there are constants γ' and γ'' such that $r'^2 \geq \gamma' r$ and $r''^2 \leq \gamma'' r$.*

If we denote by $B(r)$ a Bregman ball of radius r, we thus have $\gamma' \pi r^2 \leq \text{area}(B(r)) \leq \gamma'' \pi r^2$. When F is of class C^2 and $\mathcal{D}$ is compact, a packing argument shows that the algorithm cannot insert infinitely many points. Moreover, the size of the sample output by the algorithm is asymptotically optimal, up to a multiplicative constant. This is formally stated in the next lemma. Write $\mathcal{D}^{\leq \varepsilon} = \{\mathbf{x} \| \ \exists \mathbf{y} \in \mathcal{D}, D_F(\mathbf{x}, \mathbf{y}) \leq \varepsilon\}$.

LEMMA 7.4. (PROOF IN [18]) *The algorithm terminates. If P_t denote the final set of sample points, we have $\frac{\text{area}(\mathcal{D})}{\gamma'' \pi \varepsilon^2} \leq |P_t| \leq \frac{4\text{area}(\mathcal{D}^{\leq \frac{\varepsilon}{2}})}{\gamma' \pi \varepsilon^2}$.*

In the technical report [18], we further give related results on the geometry of Bregman divergences and its applications.

References

[1] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Oxford University Press, 2000.

[2] F. Aurenhammer, *Power diagrams: Properties, algorithms and applications*, SIAM Journal of Computing, 16 (1987), pp. 78–96.

[3] F. Aurenhammer and R. Klein, *Voronoi Diagrams*, Handbook of Computational Geometry, J. Sack and G. Urrutia, Eds., Elsevier, 2000, pp. 201–290.

[4] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, *Clustering with Bregman divergences*, Journal of Machine Learning Research , 6 (2005), pp. 1705–1749.

[5] J.-D. Boissonnat and M. Karavelas, *On the combinatorial complexity of Euclidean Voronoi cells and convex hulls of d-dimensional spheres*, Proc. 14th ACM-SIAM Sympos. Discrete Algorithms , 2003, pp. 305–312.

[6] J.-D. Boissonnat, C. Wormser, and M. Yvinec, *Anisotropic diagrams: Labelle Shewchuk approach revisited*, 17th Canadian Conference on Computational Geometry, 2005, pp. 266–269.

[7] J.-D. Boissonnat and M. Yvinec, *Algorithmic Geometry*, Cambridge University Press, 1998.

[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, 7 (1967), pp. 200–217.

[9] K. Q. Brown, *Voronoi diagrams from convex hulls*, Inf. Process. Lett., 9 (1979), pp. 223–228.

[10] B. Chazelle, *An optimal convex hull algorithm in any fixed dimension*, Discrete Computational Geometry, 10 (1993), pp. 377–409.

[11] Q. Du , V. Faber and M. Gunzburger, *Centroidal Voronoi Tesselations : Applications and Algorithms*, SIAM Review, 41 (1999), pp. 637–676.

[12] Y. Eldar, M. Lindenbaum, M. Porat and Y. Y. Zeevi, *The Farthest Point Strategy for Progressive Image Sampling*, IEEE Trans. on Image Processing, 6(9) (1997), pp. 1305–1315.

[13] M. Herbster and M. Warmuth, *Tracking the best regressor*, Proc. 11th Conference on Computational Learning Theory, 1998, pp. 24–31.

[14] R. Klein, *Concrete and Abstract Voronoi Diagrams*, vol. 400 of Lecture Notes in Computer Science, Springer, 1989. ISBN 3-540-52055-4.

[15] G. Leibon and D. Letscher, *Delaunay triangulations and Voronoi diagrams for Riemannian manifolds*, Proc. 16th Symposium on Computational Geometry, 2000, pp. 341–349.

[16] S. P. Lloyd, *Least squares quantization in PCM*, IEEE Trans. Inform. Theory (1982), pp. 129–137

[17] P. McMullen, *The maximum numbers of faces of a convex polytope*, J. Combinatorial Theory, Ser. B, 10 (1971), pp. 179–184.

[18] F. Nielsen, J.-D. Boissonnat, and R. Nock, *On Bregman Voronoi diagrams*, INRIA technical report, 2006.

[19] R. Nock and F. Nielsen, *Fitting the smallest enclosing Bregman ball*, 16th European Conference on Machine Learning, 2005, pp. 649–656.

[20] K. Onishi and H. Imai, *Voronoi diagram in statistical parametric space by Kullback-Leibler divergence*, in Proc. 13th Symposium on Computational Geometry , 1997, pp. 463–465.

[21] K. Onishi and N. Takayama, *Construction of Voronoi diagram on the upper half-plane*, IEICE Transactions, 79-A (1996), pp. 533–539.

[22] V. T. Rajan, *Optimality of the Delaunay triangulation in $\mathbb{R}^d$*, Disc. & Comp. Geom., 12 (1994), pp. 189–202.

[23] J. Ruppert, *A Delaunay Refinement Algorithm for Quality 2-Dimensional Mesh Generation*, J. Algorithms, 18 (1995), pp. 548–585.

[24] K. Sadakane, H. Imai, K. Onishi, M. Inaba, F. Takeuchi, and K. Imai, *Voronoi diagrams by divergences with additive weights*, Proc. 14th Symposium on Computational Geometry, 1998, pp. 403–404.

[25] M. Sharir, *Almost tight upper bounds for lower envelopes in higher dimensions*, Discrete Comput. Geom., 12 (1994), pp. 327–345.